



## Evaluating the value of a network of cosmic-ray probes for improving land surface modelling

Roland Baatz<sup>1,2</sup>, Harrie-Jan Hendricks Franssen<sup>1,2</sup>, Xujun Han<sup>1,2</sup>, Tim Hoar<sup>3</sup>, Heye R. Bogaen<sup>1</sup> and Harry Vereecken<sup>1,2</sup>

<sup>1</sup>Agrosphere (IBG-3), Forschungszentrum Jülich GmbH, 52425 Jülich, Germany.

<sup>2</sup>HPSC-TerrSys, 52425 Jülich, Germany.

<sup>3</sup>NCAR Data Assimilation Research Section, Boulder, CO, USA.

*Correspondence to:* Roland Baatz (r.baatz@fz-juelich.de)



**Abstract:** Land surface **models can model** matter and energy fluxes between the land surface and atmosphere, and provide a lower boundary condition to atmospheric circulation models. For these applications, accurate soil moisture quantification is highly desirable but not always possible given **limited observations and limited** subsurface data accuracy. **Cosmic-ray probes (CRPs) offer an interesting alternative to indirectly measure soil moisture and provide an observation that can be assimilated**  
 5 **into land surface models for improved soil moisture prediction.** Synthetic studies have shown the potential to estimate subsurface parameters of land surface models with the assimilation of CRP observations. In this study, the potential of a network of CRPs for estimating subsurface parameters and improved soil moisture states is tested in a real-world case scenario using the local ensemble transform Kalman filter with the Community Land Model. The potential of the CRP network was tested by assimilating CRP-data for the years 2011 and 2012 (with or without soil hydraulic parameter  
 10 estimation), followed by the verification year 2013. This was done using (i) the regional soil map as input information for the simulations, and (ii) an erroneous, biased soil map. For the regional soil map, soil moisture characterization was only improved in the assimilation period but not in the verification period. For the biased soil map, soil moisture characterization improved in both periods strongly from a  **$E_{RMS}$**  of  $0.11 \text{ cm}^3/\text{cm}^3$  to  $0.03 \text{ cm}^3/\text{cm}^3$  (assimilation period) and from  $0.12 \text{ cm}^3/\text{cm}^3$  to  $0.05 \text{ cm}^3/\text{cm}^3$  (verification period) and the estimated soil hydraulic parameters were after assimilation closer to the ones of  
 15 the regional soil map. Finally, the value of the CRP network was also evaluated with jackknifing data assimilation experiments. It was found that the CRP network is able to improve soil moisture estimates at locations between the assimilation sites from a  $E_{RMS}$  of  $0.12 \text{ cm}^3/\text{cm}^3$  to  $0.06 \text{ cm}^3/\text{cm}^3$  (verification period), but again only if the initial soil map was biased.

## 1 Introduction

Soil water content (SWC) is a key variable of land surface hydrology and has a strong control on the partitioning of net radiation between latent and sensible heat flux. Knowledge of SWC is relevant for the assessment of plant water stress and agricultural production, as well as runoff generation as a response to precipitation events (Vereecken et al., 2008; Robinson et al., 2008). In atmospheric circulation models, SWC is important as a lower boundary condition and is calculated as a state  
 25 variable in land surface models. Coupling of atmospheric circulation models and land surface models allows quantifying the role of soil moisture on atmospheric processes such as soil moisture-precipitation feedbacks (Koster et al., 2004; Eltahir, 1998) and summer climate variability and drought (Oglesby and Erickson, 1989; Sheffield and Wood, 2008; Seneviratne et al., 2006; Bell et al., 2015). It is therefore important to improve the modelling and prediction of SWC, but this is hampered by model deficiencies and lack of high quality data (Vereecken et al., 2016). Soil moisture measured by **space-born** remote  
 30 sensing technologies provides information over large areas (e.g. **Temimi et al., 2014**). However, space-born remote sensing supplies only information on the upper few centimeters, and data are not reliable for areas with dense vegetation. Therefore, in this paper an alternative source for soil moisture information is explored. Cosmic-ray probes (CRPs) measure fast neutron intensity which allows estimating SWC at an **intermediate scale (Zreda et al., 2008; Desilets et al., 2010; Cosh et al., 2016; Lv et al., 2014)** which is closer to the desired application scale of land surface models (Ajami et al., 2014; Chen et al.,



2007;Shrestha et al., 2014). **Fast neutrons originate from moderation of secondary cosmic particles from outer space by terrestrial atoms.** These particles are mainly fast neutrons, which are moderated most effectively by hydrogen **because of the similar atomic mass.** Therefore, the corresponding neutron intensity measured by CRPs strongly depends on the amount of hydrogen within the CRP footprint, allowing for a continuous non-invasive soil moisture estimate over an area of ~15 ha (Kohli et al., 2015). The spatial ~~extend~~ of this measurement is desirable as it matches with the desired grid cell size of a land surface model (Crow et al., 2012) and small scale heterogeneities are averaged over a larger area (Franz et al., 2013a;Desilets and Zreda, 2013). Vertical measurement depth ranges from a maximum of ~70 cm under completely dry conditions and decreases to roughly ~12 cm under wet conditions (e.g. 40 vol. % soil moisture) (Bogena et al., 2013). Worldwide several CRP networks exist, like the North American **COSMOS network (Zreda et al., 2012), the German CRP network (Baatzt et al., 2014) installed in the context of the TERENO infrastructure measure (Zacharias et al., 2011) and the Australian COSMoZ network (Hawdon et al., 2014).**

**Soil moisture data assimilation** provides a way to improve imperfect land surface model predictions with measured soil moisture data by merging model predictions and data, and can consider the uncertainty of initial conditions, model parameters and model forcings. Ensemble Kalman Filtering (EnKF) is one of the most commonly applied data assimilation methods (Evensen, 1994;Burgers et al., 1998). ~~Soil moisture data assimilation has been the subject of intensive study for more than a decade now. An early contribution was provided by Houser et al. (1998) who assimilated remotely sensed soil moisture observations from a microwave radiometer into a land atmosphere transfer scheme using four-dimensional variational data assimilation. Rhodin et al. (1999) assimilated soil moisture data for a four day period in order to obtain an improved characterization of the lower boundary condition for an atmospheric circulation model. They also used a variational data assimilation approach.~~ More recently, the Ensemble Kalman Filter (Reichle et al., 2002a;Dunne and Entekhabi, 2005;Crow, 2003), the Extended Kalman Filter (Draper et al., 2009;Reichle et al., 2002b), four-dimensional variational methods (Hurkmans et al., 2006) and the Local Ensemble Transform Kalman Filter (Han et al., 2015;Han et al., 2013) were applied for updating soil moisture states in land surface models. Reichle et al. (2002a) performed a synthetic experiment using L-band microwave observations of the Southern Great Plains Hydrology Experiment (Jackson et al., 1999) to analyze the effect of ensemble size and prediction error. Dunne and Entekhabi (2005) showed that an Ensemble Kalman Smoother approach, where data from multiple time steps was assimilated to update current and past states, can yield a reduced prediction error compared to a pure filtering approach.

More recent work addressed joint state-parameter estimation in hydrologic land surface models with data assimilation methods. **Joint state-parameter estimation with EnKF is possible** by an augmented state vector approach (Chen and Zhang, 2006), a dual approach (Moradkhani et al., 2005) or an approach with an additional external optimization loop (Vrugt et al., 2005). Pauwels et al. (2009) optimized soil hydraulic parameters of a land surface model with synthetic aperture radar data. Lee (2014) used Synthetic Aperture Radar soil moisture data to estimate soil hydraulic properties at the Tibetan plateau



using the EnKF and a Soil Vegetation Atmosphere Transfer model. Bateni and Entekhabi (2012) assimilated land surface temperature with an Ensemble Kalman Smoother and achieved a better estimate of the partitioning of energy between sensible and latent heat fluxes. Han et al. (2014b) updated soil hydraulic parameters of the Community Land Model (CLM) by assimilation of synthetic brightness temperature data with the Local Ensemble Transform Kalman Filter (LETKF) (Hunt et al., 2007). Shi et al. (2014) used the Ensemble Kalman Filter for a synthetic multivariate data assimilation problem with a land surface model and then applied it to real data (Shi et al., 2015). The cases illustrated a way to use real world data for estimating several parameters in hydrologic land models. (Kurtz et al., 2016) developed a particular CPU-efficient data assimilation framework for the coupled land surface-subsurface model TerrSysMP (Shrestha et al., 2014). They successfully updated  $2 \times 10^7$  states and parameters in a synthetic experiment. Whereas these studies were made with land surface models, also in soil hydrological applications recently data assimilation was used to estimate soil hydraulic parameters. Early work was by Wu and Margulis (2011, 2013) in the context of real-time control of waste water reuse in irrigation. Erdal et al. (2014) investigated the role of bias in the conceptual soil model and explored bias aware EnKF as a way to deal with it. Erdal et al. (2015) focused on handling of strong **non-Gaussianity** of the state variable in EnKF under very dry conditions. Montzka et al. (2013;2011) explored the role of the particle filter for handling non-Gaussianity in soil hydrology data assimilation. They showed that the ability of a data assimilation system to correct the soil moisture state and estimate hydraulic parameters strongly depends on the nonlinear character of the soil moisture retention characteristic. Song et al. (2014) worked on a modified iterative filter to handle the non-linearity and non-Gaussianity of data assimilation for the vadose zone. For a further literature review on data assimilation in the context of hydrological and land surface models we refer to Reichle (2008) and Montzka et al. (2012).

Shuttleworth et al. (2013) developed the Cosmic Ray Soil Moisture Interaction Code (COSMIC), which is a forward operator to be applied for assimilating neutron intensity observations from CRPs. The COSMIC code was evaluated for several sites (Baatz et al., 2014;Rosolem et al., 2014). Its capability to propagate surface soil moisture information into the deeper soil column was analyzed by Rosolem et al. (2014). The COSMIC operator was successfully implemented in the Data Assimilation Research Testbed (Rosolem et al., 2014) to allow for state updating by the Ensemble Adjustment Kalman Filter (Anderson, 2001). The COSMIC operator was implemented in a python interface that couples the land surface model CLM and the LETKF for joint state parameter updating (Han et al., 2015). Neutron counts measured by CRP have been used in data assimilation studies to update model states (Han et al., 2015;Rosolem et al., 2014). Soil hydraulic parameters were also updated by assimilation of neutron counts (Han et al., 2016), but only for a synthetic study which showed its feasibility. CRPs were also used for inverse estimation of soil hydraulic parameters of the Hydrus-1D model (Villarreyes et al., 2014).

This work further explores the value of measured neutron intensity by CRPs to improve modelling of terrestrial systems at the catchment scale (Simmer et al., 2015) using a land surface model. ~~Compared to existing work~~ the main novelties are:



(i) Data from a network of nine CRPs were assimilated in the Community Land Model version 4.5 (CLM) with an evaluation of the information gain by this assimilation at the larger catchment scale. Until now evaluations with CRPs were made for a single location, but not for a complete network of CRPs. It is a very important question whether CRPs can also improve the soil moisture characterization at the larger catchment scale and how dense the CRP network should be. The high variability of soil moisture at a short distance could potentially limit the CRP measurement value and make updating of soil moisture contents further away from the sensor meaningless. On the other hand, soil maps and atmospheric forcings show spatial correlations over larger distances which suggests that CRP measurements potentially carry important information to update soil moisture contents for larger regions. If it is found that CRP networks with a density like in this study (10 stations per 2354 km<sup>2</sup>) can improve soil moisture content characterization at the larger catchment scale, this is of high relevance and importance for agricultural applications, flood prediction and protection, and regional weather prediction (Whan et al., 2015; Koster et al., 2004; Seneviratne et al., 2010). The main research question addressed in this paper is therefore whether a CRP network of the density in this study can improve large scale soil moisture characterization by state and parameter updates.

(ii) Soil hydraulic parameters were updated together with the soil moisture states in a real-world case study at the larger catchment scale. The study in this paper also allows some evaluation of the feasibility of the updated large scale soil hydraulic parameters.

In the following paragraphs are presented the model site and the measurements (2.1), the Community land Model and its parameterization (2.2), the COMIC forward model (2.3) and the data assimilation procedure (2.4).

## 2 Materials and methods

### 2.1 Site description and measurements

The model domain, the Rur catchment (2354 km<sup>2</sup>), is situated in western Germany and illustrated in Fig. 1. Most prominent vegetation types are agricultural land use (mainly in the North), grassland, and coniferous and deciduous forest. The altitude varies between 15 m a.s.l. in the flat northern part and 690 m a.s.l. in the hilly southern part. Precipitation, evapotranspiration and land use follow the topography. Annual precipitation ranges between less than 600 mm in the North to 1200 mm in the hilly South (Montzka et al., 2008). Annual potential evapotranspiration varies between 500 mm in the South and 700 mm in the North (Bogena et al., 2005). The Rur catchment CRP network comprises nine CRPs (CRS1000, HydroInnova LLC, 2009) which were installed in 2011 and 2012 (Baatz et al., 2014). Climate and soil texture of the CRP sites can be found in Table 1. The CRPs were calibrated in the field using gravimetric soil samples. At each site, 18 soil samples were taken along three circles with distances of 25, 75 and 175 meters from the CRP, six samples evenly distributed along each circle. Each sample was extracted with a 50.8 x 300 mm round HUMAX soil corer (Martin Burch AG, Switzerland). The samples were split into 6 sub-samples with 5 cm length each and oven dried at 105 °C for 48 hours to measure dry soil bulk density and soil moisture. Lattice water was determined for each site using a heat conductivity detector. Soil bulk density, soil moisture,



lattice water and 12 hour averaged measured neutron intensity were used to determine calibration parameters specific for each CRP and the COSMIC operator.

## 2.2 Community land model and parameterization

5 The Community Land Model version 4.5 (CLM) was the land surface model of choice for simulating water and energy exchange between the land surface and the atmosphere (Oleson et al., 2013). Some of the key processes which are solved by CLM are radiative transfer in the canopy space, interception of precipitation by the vegetation and evaporation from intercepted water, water uptake by vegetation and transpiration, soil evaporation, photosynthesis, as well as water and energy flow in the subsurface. SWC in CLM is influenced by precipitation, infiltration into the soil, water uptake by vegetation, surface evaporation and surface and subsurface runoff. Oleson et al. (2013) provide further details on CLM4.5. To limit the scope and complexity of this study, CLM was run using satellite phenology, e.g. prescribed leaf area index data and the biogeochemical module turned off.

The spatial domain is discretized by rectangular grid cells by CLM. Each grid cell may have several types of land cover: Lake, urban, vegetated, wetland, and glacier. The vegetated part of the grid cell can be covered by several plant functional types which are all linked to a single soil column. The soil column is vertically discretized by ten soil layers and five bedrock layers. Layer thickness increases exponentially from 0.007 m at the surface to 2.86 m for layer 10. Vertical water flow in soils is modelled by the 1D Richards equation. Soil hydraulic parameters are determined from sand and clay content using pedotransfer functions for the mineral soil fraction (Clapp and Hornberger, 1978; Cosby et al., 1984), and organic matter content for the organic soil fraction (Lawrence and Slater, 2008).

The following equations describe how soil texture and organic matter define the soil hydraulic properties in CLM such as porosity, hydraulic conductivity, the empirical exponent  $B$  and soil matric potential. Hydraulic conductivity ( $k[z]$  in mm/s) at the depth  $z$  between two layers ( $i$  and  $i+1$ ) is a function of soil moisture ( $\theta$  in  $\text{m}^3/\text{m}^3$  in layers  $i$  and  $i+1$ ), saturated hydraulic conductivity ( $k_{sat}$  in mm/s at  $z$ ), saturated soil moisture ( $\theta_{sat}$  in  $\text{m}^3/\text{m}^3$ ) and the empirical exponent  $B$ :

$$k[z] = \begin{cases} \phi_{ice} k_{sat,z} \left[ \frac{0.5(\theta_i + \theta_{i+1})}{0.5(\theta_{sat,i} + \theta_{sat,i+1})} \right]^{2B_i+3} & 1 \leq i \leq N_{levsoi} - 1 \\ \phi_{ice} k_{sat,z} \left( \frac{\theta_i}{\theta_{sat,i}} \right)^{2B_i+3} & i = N_{levsoi} \end{cases} \quad 1$$

where  $\phi_{ice}$  is the ice impedance factor. The ice impedance factor was implemented to simplify an increased tortuosity of water flow in a partly frozen pore space. It is calculated with  $\phi_{ice} = 10^{-\Omega F_{ice}}$  using the resistance factor  $\Omega = 6$  and the frozen fraction of soil porosity  $F_{ice} = \theta_{ice}/\theta_{sat,i}$ . Soil hydraulic properties are calculated separately for the mineral ( $min$ ) and organic matter ( $om$ ) soil components. Total porosity  $\theta_{sat,i}$  is calculated using the fraction of organic matter ( $f_{om,i}$ ) with:



$$\theta_{sat,i} = (1 - f_{om,i})\theta_{sat,min,i} + f_{om,i}\theta_{sat,om} \quad 2$$

where the organic matter porosity is  $\theta_{sat,om} = 0.9$  and sand content in % determines the mineral soil porosity  $\theta_{sat,min}$  as:  
 $\theta_{sat,min} = 0.489 - 0.00126 \times \%sand$  3

Analogous, the exponent  $B$  is calculated with

$$B_i = (1 - f_{om,i})B_{min,i} + f_{om,i}B_{om} \quad 4$$

Where  $B_{om} = 2.7$  is the organic exponent and the mineral exponent  $B_{min,i}$  is determined by clay content in % with:  
 $B_{min,i} = 2.91 + 0.159 \times \%clay$  5

Saturated hydraulic conductivity is calculated for a connected and an unconnected fraction of the grid cell with:

$$k_{sat}[z_i] = (1 - f_{perc})k_{sat,uncon}[z_i] + f_{perc,i}k_{sat,om}[z_i] \quad 6$$

- 5 where  $f_{perc,i}$  is the fraction of a grid cell where water flows with saturated hydraulic conductivity of the organic matter ( $k_{sat,om}[z_i]$  in mm/s) through the organic material only, the so called connected flow pathway, whereas the saturated hydraulic conductivity of the unconnected part ( $k_{sat,uncon}[z_i]$  in mm/s) depends on organic and mineral saturated soil hydraulic conductivity:

$$k_{sat,uncon} = (1 - f_{perc}) \left( \frac{1 - f_{om}}{k_{sat,min}} + \frac{f_{om} - f_{perc}}{k_{sat,om}} \right)^{-1} \quad 7$$

where saturated hydraulic conductivity for mineral soil is calculated from the grid cell sand content as:

$$k_{sat,min}[z_i] = 0.0070556 \times 10^{-0.884 + 0.0153 \times \%sand} \quad 8$$

- 10 The fraction  $f_{perc}$  is calculated with:

$$f_{perc} = 0.908 \times (f_{om} - 0.5)^{0.139} \quad f_{om} \geq 0.5 \quad 9$$

$$f_{perc} = 0 \quad f_{om} < 0.5 \quad 10$$

Soil matric potential (mm) is defined as function of saturated soil matric potential (mm) with:

$$\psi_i = \psi_{sat,i} \left( \frac{\theta_i}{\theta_{sat,i}} \right)^{-B_i} = [(1 - f_{om,i})\psi_{sat,min,i} + f_{om,i}\psi_{sat,om}] \left( \frac{\theta_i}{\theta_{sat,i}} \right)^{-B_i} \quad 11$$

where saturated organic matter matric potential is  $\psi_{sat,om} = -10.3$  mm and saturated mineral soil matric potential is calculated from sand content as:

$$\psi_{sat,min,i} = -10.0 \times 10^{1.88 - 0.0131 \times \%sand} \quad 12$$





### 2.3 Cosmic-ray forward model

SWC retrievals were calculated from neutron intensity observations with the **Cosmic**-ray Soil Moisture Interaction Code **COSMIC** (Shuttleworth et al., 2013) following calibration results and the procedure of Baatz et al. (2014). COSMIC parameterizes interactions between neutrons and atoms in the subsurface, relevant for soil moisture estimation. COSMIC was calibrated against the more complex Monte Carlo Neutron Particle model MCNPx (Pelowitz, 2005) and needs considerably less CPU-time than the MCNPx model. The reduced CPU-time need and the physically based parameterization make COSMIC a suitable data assimilation operator. The code was tested at multiple sites for soil moisture determination (Baatz et al., 2014; Rosolem et al., 2014) and analyzed in detail by Rosolem et al. (2014).

COSMIC assumes that ~~a number of~~ high energy neutrons enter the soil. In the soil, high energy neutrons are reduced by interaction with the soil leading to **isotropic generation** of fast neutrons **with less energy in each soil layer**. Before resurfacing, fast neutrons are **reduced** again by **soil interaction** (Shuttleworth et al., 2013). The **number of neutrons  $N_{CRP}$  that reaches the CRP** can be summarized in a single integral as

$$N_{CRP} = N_{COSMIC} \int_0^{\infty} \left\{ A(z) [\alpha \rho_s(z) + \rho_w(z)] \exp \left( - \left[ \frac{m_s(z)}{L_1} + \frac{m_w(z)}{L_2} \right] \right) \right\} \cdot dz \quad 13$$

where  $N_{COSMIC}$  is an empirical coefficient that is CRP specific and needs to be estimated by calibration,  $A(z)$  is the integrated average attenuation of fast neutrons,  $\alpha = 0.404 - 0.101 \times \rho_s$  is the site specific empirical coefficient for the creation of fast neutrons by soil,  $\rho_s$  is the dry soil bulk density in  $\text{g/cm}^3$ ,  $\rho_w$  is the total soil water density in  $\text{g/cm}^3$ ,  $m_s$  and  $m_w$  are the mass of soil and water, respectively, per area in  $\text{g/cm}^2$ .  $L_1 = 162.0 \text{ g cm}^{-2}$  and  $L_2 = 129.1 \text{ g cm}^{-2}$  are **empirical coefficients** that were estimated using the MCNPx code (Shuttleworth et al., 2013). The integrated average attenuation of fast neutrons  $A(z)$  can be found numerically by solving

$$A(z) = \left( \frac{2}{\pi} \right) \int_0^{\pi/2} \exp \left( \frac{-1}{\cos(\theta)} \left[ \frac{m_s(z)}{L_3} + \frac{m_w(z)}{L_4} \right] \right) \cdot d\theta \quad 14$$

where  $\theta$  is the angle along a vertical line below the CRP detector to the element that contributes to the attenuation of fast neutrons,  $L_3 = -31.65 + 99.29 \times \rho_s$  is determined from soil bulk density and  $L_4 = 3.16 \text{ g cm}^{-2}$  is another **empirical coefficient** estimated using the MCNPx code (Shuttleworth et al., 2013). The COSMIC operator is discretized into 300 **vertical layers** of one cm thickness up to a depth of three meters. For each CLM grid cell in the model domain, simulated SWC is used to generate a SWC retrieval using the COSMIC code. Simulated SWC is handed from the CLM simulation history files to the COSMIC operator. Given the vertical SWC distribution of the individual CLM soil column, COSMIC internally calculates the contribution of each layer to the simulated neutron intensity signal at the COSMIC soil surface. In this study, the contribution of each layer was used to calculate the weighted CLM SWC retrieval corresponding to the vertical distribution of simulated SWC in each grid cell. Measured neutron intensity of CRPs was used to inversely





determine a CRP SWC retrieval as by Baatz et al. (2014) assuming a homogeneous vertical SWC distribution. Then, the weighted CLM SWC retrieval is used in the data assimilation scheme to relate the CRP SWC retrieval to the model state.

## 2.4 Data assimilation

5 This study uses the local ensemble transform Kalman filter (LETKF) (Hunt et al., 2007) to assimilate SWC retrievals by CRPs into the land surface model CLM. Besides other Ensemble Kalman Filter variants, the LETKF is applied in atmospheric sciences (Liu et al., 2012; Miyoshi and Kunii, 2012), ocean science (Penny et al., 2013) and also in land surface hydrology (Han et al., 2014a; Han et al., 2015). Updates were calculated either for states or jointly for states and parameters. For state updates only, the LETKF was used as proposed by Hunt et al. (2007). Calculations were made for an ensemble of  
 10 model simulations which differed related to variations in model forcings and input parameters. The states of the different ensemble members are indicated by  $\mathbf{x}_i^f$  where  $i=1, \dots, N$  and  $N$  is the number of ensemble members. The individual state vectors  $\mathbf{x}_i^f$  contain the CLM-simulated SWC of the ten soil layers and the vertically weighted SWC retrieval obtained with the COSMIC operator. For each grid cell, a vector  $\mathbf{X}^f$  can be constructed which contains the deviations of the simulated states with respect to the ensemble mean  $\bar{\mathbf{x}}^f$ :

$$\mathbf{X}^f = [\mathbf{x}_1^f - \bar{\mathbf{x}}^f, \dots, \mathbf{x}_N^f - \bar{\mathbf{x}}^f] \quad 15$$

15 In case of joint state-parameter updates, a state augmentation approach was followed (Hendricks Franssen and Kinzelbach, 2008; Han et al., 2014b). In this case, the augmented model state vector  $\mathbf{X}^f$  is constructed from the weighted SWC, and the grid cell's sand, clay and organic matter content.

In order to relate the measured neutron intensity with the simulated SWC of CLM, the observation operator  $\mathbf{H}$  (COSMIC) is  
 20 applied on the measured neutron intensity in order to obtain the expected weighted SWC retrieval for each of the stochastic realizations:

$$\mathbf{y}_i^f = \mathbf{H}(\mathbf{x}_i^f) \quad 16$$

The ensemble realizations of the modelled SWC retrieval  $\mathbf{y}_1^f$  to  $\mathbf{y}_N^f$  with respect to the ensemble mean  $\bar{\mathbf{y}}^f$  are stored in the vector  $\mathbf{Y}^f$ :

$$\mathbf{Y}^f = [\mathbf{y}_1^f - \bar{\mathbf{y}}^f, \dots, \mathbf{y}_N^f - \bar{\mathbf{y}}^f] \quad 17$$

The observation error correlation was reduced in space by the factor  $f_{red}$  using the spherical model:

$$f_{red} = 1 - (1.5 \times d/d_{max}) + (0.5 \times [d/d_{max}]^3) \quad 18$$

25 where  $d$  is the distance to the observation and  $d_{max} = 40km$  is the maximum observation correlation length, about half the size of the catchment. Only SWC retrievals within the maximum observation correlation length were used for assimilation.



This leads to a ‘localized’ size of  $\mathbf{Y}^f$  and the observation error covariance matrix  $\mathbf{R}$ . The intermediate covariance matrix  $\mathbf{P}^a$  (also called analysis error covariance matrix) is calculated according to:

$$\mathbf{P}^a = [(N - 1)\mathbf{I} + \mathbf{Y}^{fT}\mathbf{R}^{-1}\mathbf{Y}^f] \quad 19$$

In addition, the mean weight vector  $\bar{\mathbf{w}}^a$  is obtained as follows:

$$\bar{\mathbf{w}}^a = \mathbf{P}^a \mathbf{Y}^{fT} \mathbf{R}^{-1} (\mathbf{y}^0 - \bar{\mathbf{y}}^f) \quad 20$$

where  $\mathbf{y}^0$  is CRP SWC retrieval. In the ensemble space, a perturbation matrix  $\mathbf{W}^a$  is calculated from the symmetric square root of  $\mathbf{P}^a$ :

$$\mathbf{W}^a = [(N - 1)\mathbf{P}^a]^{1/2} \quad 21$$

The final analysis  $\mathbf{X}^a$  is obtained from:

$$\mathbf{X}^a = \bar{\mathbf{x}}^f + \mathbf{X}^f [\bar{\mathbf{w}}^a + \mathbf{W}^a] \quad 22$$

A more detailed description of the LETKF can be found in (Hunt et al., 2007) and details on the implementation of the LETKF in combination with CLM are given by (Han et al., 2015).

## 10 3 Model and Experiment Setup

### 3.1 Model Setup

In this study, discretization and parameterization of the hydrological catchment was done on the basis of high resolution data. The Rur catchment domain is spatially discretized by rectangular grid cells of 0.008 degree size (~750 m). The model time step was set to hourly. Land cover was assumed to consist of vegetated land units only, and a single plant functional type (PFT) for each grid cell was defined. The plant functional types were derived from a remotely sensed land use map using RapidEye and ASTER data with 15 m resolution (Waldhoff, 2012). Sand content, clay content and organic matter content were derived from the high resolution regional soil map BK50 (Geologischer Dienst Nordrhein-Westfalen, 2009). The BK50 soil map provides the high resolution soil texture for the catchment and is the most detailed soil map available for the defined region. As an alternative, simulations were also performed for a biased soil texture distribution with a fixed sand content of 80 % and clay content of 10 % (S80 soil map). This represents a large error with respect to the expected true soil properties. It allows evaluating the joint state-parameter estimation approach because given the expected bias, we can evaluate whether and to what extent the soil properties are modified by the data assimilation to be closer to the available high resolution soil map. In addition, in many regions across the Earth a high resolution soil map is not available and land surface models which are applied for those regions, for example in the context of global simulations, might be strongly affected by the error in soil properties. It was tested how this impacted the simulation results. Maximum saturated fraction, a surface parameter which is used for runoff generation, was calculated from a 10 meter digital elevation model (scilands GmbH, 2010). Leaf area index data were derived from monthly averaged Moderate Resolution Imaging Spectrometer data



(MODIS). CLM was supplied with hourly atmospheric forcing data from a reanalysis data set for the years 2010 to 2013 from the German Weather Service (DWD). The data was downscaled from a resolution of  $2.8 \text{ km}^2$  to the CLM resolution using linear interpolation based on Delaunay triangulation. Forcing data include precipitation in mm/s, incident solar and longwave radiation in  $\text{W/m}^2$ , air temperature in K, air pressure in hPa, wind speed in m/s and relative humidity in kg/kg at the lowest atmospheric level.

### 3.2 Model ensemble

Uncertainty was introduced into the regional CLM model by perturbed soil parameters and forcings. Contents of sand, clay and organic matter were perturbed with spatially correlated noise from a uniform sampling distribution with mean zero and standard deviation 10 % or 30 % (Han et al., 2015). By perturbing texture, soil parameters are also perturbed through the pedotransfer functions used in CLM as specified in Sect. 2.2. Precipitation ( $\sigma = 0.5$  or  $1.0$ ; lognormal distribution) and shortwave radiation ( $\sigma = 0.3$ ; lognormal distribution) were perturbed with multiplicative noise with mean equal to one. Longwave radiation ( $\sigma = 20 \text{ W m}^{-2}$ ) and air temperature ( $\sigma = 1\text{K}$ ) were perturbed with additive noise. The forcing perturbations were imposed with correlations in space (5 km) using a fast Fourier transform. Correlation in time was introduced with an AR(1)-model with autoregressive parameter=0.33. These correlations and standard deviations were chosen based on previous data assimilation experiments (Reichle et al., 2010; Kumar et al., 2012; De Lannoy et al., 2012; Han et al., 2015). In this work, only results for precipitation perturbation with  $\sigma = 0.5$  will be shown as results for  $\sigma = 1.0$  were very similar. An ensemble size of 95 realizations was used in the simulations. Based on previous work (Baatz et al., 2015), the SWC retrieval uncertainty for CRPs was estimated to be  $0.03 \text{ cm}^3/\text{cm}^3$ .

### 3.3 Experiment set-up

All simulation experiments in this study used initial conditions from a single five year spin-up run. For the five year spin-up run, a single forcing data set of the year 2010 was repeatedly used as atmospheric input. The soil moisture regime became stable after five years spin-up period, and additional spin-up simulations would not affect soil moisture in the consecutive years. After this five year spin-up, soil parameters and forcing data of the consecutive years were perturbed. From 1<sup>st</sup> Jan. 2011 onwards, CLM was propagated forward with an ensemble of 95 realizations. On 20<sup>th</sup> Mar. 2011, the first SWC retrieval was assimilated and assimilation of SWC retrievals continued until 31<sup>st</sup> Dec. 2012. From 1<sup>st</sup> January 2013 to 31<sup>st</sup> December 2013 the model was propagated forward without data assimilation but with an ensemble of 95 realizations. The year 2013 was used exclusively as evaluation period for data assimilation experiments.

In total, 26 simulation experiments were carried out using different setups (Table 2). Two open loop simulations were run for the BK50 soil map (OL-BK50) and the S80 soil map (OL-S80), respectively, without data assimilation and soil parameter perturbation of 30 %. These simulations are referred to as reference runs for the respective soil map. Simulation results of data assimilation runs were compared to the reference runs for quantification of data assimilation benefits. Simulations were



done with joint state-parameter estimation (PAR-), two for the S80 soil map (PAR-S80-) and two for the BK50 soil map (PAR-BK50-), for which soil texture was perturbed by 10 % and 30 %. Two simulations were done with state updates only for the BK50 soil map (Stt-BK50) and the S80 soil map (Stt-BK50), where soil texture was perturbed by 30 %. These eight simulations form the basic set of experiments.

5

Besides the data assimilation experiments also a larger number of jackknifing simulations was run to evaluate the data assimilation performance. These simulations allow evaluating the impact of the CRP network to improve SWC characterization at other locations, without CRP. In a jackknife experiment, data from eight CRP locations were assimilated and one CRP was excluded from the assimilation for evaluation purpose. At the evaluation location, simulated SWC (which is affected by the assimilation of the other eight probes) was compared to CRP SWC retrievals. For jackknife simulations, the perturbation of soil texture was set to 30 % and precipitation perturbation was done with  $\sigma = 0.5$ . States and parameters at these sites were jointly updated, and simulations were made using the BK50 and the S80 soil maps as input. Therefore, a total of 18 jackknife simulations (jk-S80-\* and jk-BK50-\*) was performed (two soil maps times nine different simulations leaving away one CRP at a time).

10

15

Simulation results were evaluated with the root mean square error ( $E_{RMS}$ ):

$$RMSE = \sqrt{\frac{\sum_{t=1}^n (SWC_{t,CLM} - SWC_{t,CRP})^2}{n}} \quad 23$$

where  $n$  is the total number of time steps,  $SWC_{t,CLM}$  is the CLM SWC retrieval at time step  $t$  and  $SWC_{t,CRP}$  is the CRP SWC retrieval at time step  $t$ . In case SWC was assimilated at the corresponding time step,  $SWC_{t,CLM}$  is SWC prior to assimilation. In the case the  $E_{RMS}$  is estimated at a single point in time over all CRPs available, the number of time steps  $n$  can be replaced by the number of CRPs available. The second evaluation measurement in this study is the bias:

20

$$bias = \frac{\sum_{t=1}^n (SWC_{t,CLM} - SWC_{t,CRP})}{n} \quad 24$$

## 4 Results

### 4.1 General Results

Table 3 summarizes the performance statistics in terms of  $E_{RMS}$  for the assimilation period (2011 and 2012). Presented are results for the open loop scenario and six data assimilation scenarios. Errors of open loop simulations are higher for the S80-simulation than for the BK50-simulation at all sites but Merzenhausen. At Merzenhausen  $E_{RMS}$  was  $0.054 \text{ cm}^3/\text{cm}^3$  for the S80 soil map and  $0.067 \text{ cm}^3/\text{cm}^3$  for the BK50 soil map. Open loop simulations with the S80 soil map resulted in  $E_{RMS}$ -values above  $0.10 \text{ cm}^3/\text{cm}^3$  at five of nine sites. At all sites data assimilation results with the S80 soil map improved SWC compared to the open loop simulations. This was also the case for the BK50 soil map simulations at all sites but Aachen

25



where  $E_{RMS}$  was larger than for the open loop run. In general, data assimilation improved simulations more for the S80 soil map ( $E_{RMS}$  reduced by  $0.079 \text{ cm}^3/\text{cm}^3$ ) than for the BK50 soil map ( $E_{RMS}$  reduced by  $0.01 \text{ cm}^3/\text{cm}^3$ ). Room for improvement with the BK50 soil map runs was more limited because of the smaller open loop errors. Nevertheless, after state updating alone the BK50 soil map still gave smaller errors than the S80 soil map. However, joint state-parameter estimation further improved simulation results by reducing  $E_{RMS}$ -values and the parameter updating resulted in similar  $E_{RMS}$ -values for the BK50 ( $0.028 \text{ cm}^3/\text{cm}^3$ ) and S80 soil map ( $0.03 \text{ cm}^3/\text{cm}^3$ ). The  $E_{RMS}$  for simulations with 10 % and 30 % perturbation of soil texture values **did not show very different results.**

The temporal course of soil moisture in 2011 at the two sites Gevenich and Merzenhausen is shown in Fig. 2. The figures illustrates that SWC at both sites was lower with the S80 soil map than with the BK50 soil map. ~~In Gevenich and Merzenhausen,~~ mean open loop SWC in 2011 was  $0.17 \text{ cm}^3/\text{cm}^3$  for the S80 soil map at both sites and  $0.27 \text{ cm}^3/\text{cm}^3$  for the BK50 soil map at both sites. CRP measurements at Merzenhausen started in May 2011. In the data assimilation runs SWC was immediately affected at the Merzenhausen and Gevenich sites as soon as Merzenhausen CRP SWC retrievals were assimilated. The simulated SWC for the PAR-S80-30 data assimilation run increased as compared to the S80 open loop simulation. The first observation at Gevenich was recorded on July 7<sup>th</sup>, 2011. By that date, the simulated CLM SWC retrieval was already close to the CRP SWC retrieval at the Gevenich site (Fig. 2) due to SWC updates which showed to have a beneficial impact. Fig. 2 also shows that the **BK50 open loop run was close to the observed SWC at both sites**, even without data assimilation.

Fig. 3 shows the temporal course of SWC from January 2011 to December 2013 at Heinsberg and Wildenrath. Assimilation and evaluation results are shown for the case of joint state-parameter updates (PAR-S80-30), only state updates (Stt-S80), open loop (OL-S80) and CRP SWC retrievals. At Heinsberg, results show that assimilated SWC was closer to the CRP SWC retrieval when both states and parameters were updated (PAR-S80-30) than if only states were updated (Stt-S80). This is the case in the assimilation period and in the evaluation period. At the beginning of the evaluation period, the Stt-S80 simulation shows an increase in bias between modeled CLM SWC retrievals and CRP SWC retrieval within the first few days of 2013. The bias of Stt-S80 remained throughout the evaluation period. In contrast, modeled SWC during the evaluation period was close to the CRP SWC retrieval if parameters were previously updated (PAR-S80-30).

The CRP at Wildenrath started operating on May 7<sup>th</sup>, 2012. SWC retrievals at other CRPs were assimilated already from May 2011 onwards and affected SWC at Wildenrath (Fig. 3). Until May 2012, Fig. 3 shows assimilated SWC (Stt-S80 and PAR-S80-30) was higher than open loop SWC. However, no SWC retrievals were available at the Wildenrath site for comparison during this period. When SWC retrievals from the CRP at Wildenrath became available and were assimilated into the model, assimilated (Stt-S80 and PAR-S80-30) and open loop (OL-S80) SWC were close to CRP SWC retrievals. This was the case throughout the remaining assimilation and evaluation period. These results suggest that the high sand



content of the biased soil map is not far from the optimal sand content at Wildenrath. Therefore, at Wildenrath, the high sand content of both soil maps (60 % and 80 %) resulted in good modeling results already for the open loop runs. This suggests that before May 2012, simulated SWC of the open loop runs with either soil map represented more realistic SWC than assimilated SWC during this period. This will be discussed further in the discussion section.

5

## 4.2 Verification period

The year 2013 was the verification year without data assimilation.  $E_{RMS}$  values for the evaluation period 2013 are reported in Table 4. On the one hand, BK50 data assimilation runs with joint state-parameter estimation resulted in improved SWC at three out of the nine sites compared to open loop BK50-runs. For the other six sites results worsened compared to the corresponding BK50 open loop run.  $E_{RMS}$  values increased from an average of  $0.041 \text{ cm}^3/\text{cm}^3$  (OL-BK50) over all sites to  $0.047 \text{ cm}^3/\text{cm}^3$  (PAR-BK50-30). On the other hand, for the S80 soil map, all sites except Wildenrath had significantly reduced  $E_{RMS}$  values for the case of data assimilation including parameter updating compared to the S80 open loop run. For the S80 simulations, average  $E_{RMS}$  over all sites for 2013 was on average  $0.12 \text{ cm}^3/\text{cm}^3$  for the open loop run and  $0.04 \text{ cm}^3/\text{cm}^3$  for the run including data assimilation. In case only states were updated (Stt-S80 and Stt-BK50),  $E_{RMS}$  was also slightly reduced (compared to open loop runs) for the majority of sites during the evaluation period in 2013. On average, this reduction was  $0.016 \text{ cm}^3/\text{cm}^3$  for the S80 soil map (Stt-S80) and  $0.002 \text{ cm}^3/\text{cm}^3$  for the BK50 soil map (Stt-BK50). At sites, where  $E_{RMS}$  was larger for data assimilation runs with state updating (compared to open loop runs), the increase was only  $0.001 \text{ cm}^3/\text{cm}^3$ .

Bias calculated on the basis of a comparison of hourly SWC measured by CRP and simulated for 2013 is reported in Table 5. The average bias for the S80 open loop run is  $0.11 \text{ cm}^3/\text{cm}^3$  while it is  $0.02 \text{ cm}^3/\text{cm}^3$  for the BK50 open loop run. Bias of the BK50 open loop run was positive at Merzenhausen, Gevenich, Heinsberg, and Aachen, and it was negative at Rollesbroich, Kall, RurAue, and Wuestebach. Bias was zero at Wildenrath for the BK50 open loop run. Bias of the S80 open loop run was negative at all sites indicating that modeled SWC was higher than measured SWC. Joint state-parameter updates reduced the absolute bias on average to  $0.03 \text{ cm}^3/\text{cm}^3$  (PAR-S80-30) and  $0.02 \text{ cm}^3/\text{cm}^3$  (PAR-S80-10) for the S80 soil map. In case of the BK50 soil map, the bias in 2013 increased to  $0.03 \text{ cm}^3/\text{cm}^3$  by joint state-parameter updates. State updates without parameter updates reduced the biases only marginally to  $0.01 \text{ cm}^3/\text{cm}^3$  for the BK50 soil map and to  $0.09 \text{ cm}^3/\text{cm}^3$  for the S80 soil map. This indicates that state updates also can slightly improve SWC-characterization in the verification period due to improved initial conditions.

30

## 4.3 Temporal evolution of mean $E_{RMS}$

Fig. 4 shows the temporal evolution of the hourly  $E_{RMS}$  calculated for all nine CRPs.  $E_{RMS}$  was highest for the S80 open loop run and lowest for the PAR-S80-30 simulation. State updates did not improve modeled SWC as much as joint state-parameter updates improved modeled SWC. The  $E_{RMS}$  in case of Stt-S80 also falls behind the  $E_{RMS}$  of the BK50 open loop



run through most of the time. Joint state-parameter updates for the S80 soil map improved the  $E_{RMS}$  throughout most of the time compared to the open loop simulations based on the BK50 and S80 soil maps. During the assimilation period 2011-2012, the PAR-S80-30 simulation performed best out of the four simulations. During the evaluation period 2013, OL-BK50 and PAR-S80-30 performed equally well except in summer 2013 when the PAR-S80-30 simulation yielded much higher  $E_{RMS}$ -values than the BK50 open loop run.

#### 4.4 Jackknife simulations

The jackknife simulations investigated the impact of the network of CRPs for improving estimates of SWC at locations between the CRPs, outside the network. The errors shown in Table 4 refers to the two open loop simulations (for the S80 soil map and the BK50 soil map) and the 18 jackknife simulations. All simulations with the S80 soil map resulted in an improved  $E_{RMS}$  in the jackknife simulations compared to the open loop simulation, except for Wildenrath. In all cases the  $E_{RMS}$  was smaller than  $0.10 \text{ m}^3/\text{m}^3$ . Error reduction was smaller at sites where the open loop error was smaller. At sites with large open loop  $E_{RMS}$ , the assimilation could reduce the  $E_{RMS}$  by 50 % or more. In case of the BK50 soil map, the jackknife simulations resulted in  $E_{RMS}$ -values below  $0.10 \text{ m}^3/\text{m}^3$  at all sites. However, in this case only at Merzenhausen the  $E_{RMS}$  was reduced during the data assimilation period. At Wildenrath, the  $E_{RMS}$  was highest for jk-BK50 ( $0.091 \text{ m}^3/\text{m}^3$ ) and jk-S80 ( $0.095 \text{ m}^3/\text{m}^3$ ). The average absolute bias for the jackknife experiments was  $0.04 \text{ cm}^3/\text{cm}^3$  for both soil maps, BK50 and S80, in the evaluation period 2013 (Table 5). Hence, bias in the jk-S80-\* simulations improved compared to the open loop run but not in the jk-BK50-\* simulations, where bias was already small.

#### 4.5 Temporal evolution of parameters

The temporal evolution of the percentage sand content during the assimilation period for the nine CRP sites is shown in Fig. 5 for PAR-S80-30, PAR-S80-10, PAR-BK50-30, PAR-BK50-10, jk-S80-30\* and jk-BK50-30\*. Time series start on March 20<sup>th</sup>, 2011, the date of the first assimilated CRP SWC retrieval at Wuestebach. Wuestebach and sites within the influence sphere of Wuestebach (Aachen, Kall and Rollesbroich) show also a change in sand content from this date onwards. All other sites show a change in sand content in May 2012 when Rollesbroich and Merzenhausen start operating and their data is assimilated. All sites show variability in sand content over time. Wuestebach, Kall, RurAue, Rollesbroich and Heinsberg show some peaks in the time series. Merzenhausen, Aachen, Gevenich, and Wildenrath show a smoother course compared to the other sites. Sand levels approach a constant site-specific value for the sites Merzenhausen (45 %), Kall (30 %), Gevenich (41 %), RurAue (30 %), Heinsberg (42 %) and Wildenrath (62 %) with a reasonable spread amongst the experiments. The spread in estimated sand content for the sites Wuestebach, Aachen and Rollesbroich is larger, and it seems not to have stabilized at the end of the assimilation. Sand content estimates of the jackknife simulations was close to the sand content of the other data assimilation experiments with joint state-parameter estimation at the sites Merzenhausen, Gevenich, RurAue and Heinsberg. Evolution of the sand content for the jackknife simulations showed larger deviations from the sand content estimated by other data assimilation experiments for the sites Wuestebach, Kall, Aachen, Rollesbroich and Wildenrath.





The soil hydraulic parameter B and saturated hydraulic conductivity are shown in Fig. 6 and Fig. 7 for PAR-S80-30, PAR-S80-10, PAR-BK50-30, PAR-BK50-10, jk-S80-30\* and jk-BK50-30\*. Updates of soil hydraulic parameters start in March and May 2011 with the assimilation of SWC retrievals depending on the location. The B parameter increases for all simulations. Throughout the whole assimilation period B varies considerably within short time intervals. The total range of the B parameter is between 2.7 and 14 at all sites. At the sites Merzenhausen, Kall, Aachen, Gevenich and Rollesbroich, it generally ranges between 6 and 10. At Wuestebach, Heinsberg and RurAue, B ranges most of the time between 8 and 12, and at Wildenrath, B is below 8. Initial saturated hydraulic conductivity is rather high ( $k_{sat} > 0.015$  mm/s) in case of high sand content i.e. for the S80 soil map, and rather low ( $k_{sat} < 0.005$  mm/s) in case of low sand content i.e. for the BK50 soil map. In case of the S80 soil map, at all sites except Wildenrath, high initial saturated hydraulic conductivity decreases quickly by parameter updates to values below 0.01 mm/s. The initial spread in  $k_{sat}$  values amongst the simulation scenarios decreases at most sites. At Wuestebach, Merzenhausen, Aachen, Gevenich, RurAue and Heinsberg, the spread is rather small particularly at the end of the assimilation period, while at Wildenrath  $k_{sat}$  ranges from 0.005 to 0.015 for individual experiments at the end of the assimilation period. The discussion section will elaborate more on this.

#### 4.6 Latent heat and sensible heat

Latent heat flux or evapotranspiration (ET) is another important diagnostic variable of the CLM model and of importance for atmospheric models. Results of the data assimilation experiments showed that soil texture updates altered soil moisture states significantly. In Fig. 8 it is shown that joint state-parameter estimation also altered ET. Fig. 8 shows ET within the evaluation period 2013 across the whole catchment for four simulations. On the one hand, ET was similar for both open loop simulations in the South of the catchment. On the other hand, ET in the North was up to 80 mm per year lower for the S80 open loop run compared to the BK50 open loop run. Regarding open loop runs, the differences can be linked to the drier soil conditions in case OL-S80 compared OL-BK50 simulation results. For PAR-S80-10, ET increased by up to 40 mm per year in the Northern part of the catchment through data assimilation. The differences between open loop ET and data assimilation ET were larger for the S80 soil map than for the BK50 soil map. This could be related to the larger update in SWC in case of the S80 scenario compared to the BK50 scenario.

### 5 Discussion

The applied data assimilation scheme improved soil moisture characterization in the majority of simulation experiments with the regional Community Land Model (CLM). During 2011 and 2012, the biased S80 soil map gave a  $E_{RMS}$  up to  $0.17 \text{ cm}^3/\text{cm}^3$  (at Rollesbroich) in the open loop simulation which left plenty of room for improvements. The soil map BK50 led to  $E_{RMS}$ -values in open loop simulations below  $0.05 \text{ cm}^3/\text{cm}^3$  which left little room for error reduction considering the measurement error of  $0.03 \text{ cm}^3/\text{cm}^3$ . For the simulations starting with 80 % sand content, sand content was closer to the values of the BK50 soil map after joint state-parameter estimation. However, the temporal evolution of the updated soil



texture and the soil hydraulic parameters was not stable. Temporal fluctuations imply that there may be multiple or seasonal optimal parameter values. This is also supported by the findings of the temporal behavior of  $E_{RMS}$  during the evaluation period e.g. when in the dry summer 2013 the  $E_{RMS}$  peaked in the PAR-S80-30 simulation. Many possible error sources were not subject to calibration in this study but they could be crucial for an even better soil moisture and more stable soil parameter estimation. In this study we only considered uncertainty of soil parameters, but also vegetation parameters are uncertain. Also a number of other CLM-specific hydrologic parameters (e.g. decay factor for subsurface runoff and maximum subsurface drainage) strongly influence state variables in CLM and hence show also potential for optimization. Considering this uncertainty could give a better uncertainty characterization. Precipitation is an important forcing for the model calculations and its estimate could be improved. For this study, precipitation data from the COSMO\_DE re-analysis were used. A product which optimally combines gauge measurements and precipitation estimates from radar could give better precipitation estimates. This could improve the soil moisture characterization and also potentially lead to better parameter estimates. Also other error sources like the ones related to the model structure play a significant role. This should be subject of future investigation.

Evaluation simulations for 2013 led to partly improved and partly deteriorated  $E_{RMS}$  values when the BK50 soil map was used as prior information on the soil hydraulic properties. The simulations with the S80 soil map on the contrary showed an improved soil moisture characterization in all simulation scenarios and the updated soil hydraulic parameter estimates for those simulations approached the values of the BK50 soil map. These results indicate that the soil hydraulic parameters derived from the BK50 soil map were already well suited for soil moisture predictions and updating soil texture and soil parameters could not improve further the results.  $E_{RMS}$  values for simulations with state updates only (Stt-BK50 and Stt-BK50) in 2013 imply the beneficial role of state updates only. However, the improvements in the evaluation period by state updates (without parameter values) are small compared to the improvements obtained by joint state-parameter estimation. This illustrates the benefits of joint state-parameter updates compared to state updates only, and that soil moisture states are strongly determined by soil hydraulic parameters. It also illustrates that the improved characterization of soil moisture states in the assimilation period which results in improved initial states for the verification period loses its influence in the verification period fast over time.

The jackknife simulations illustrated that a network of CRPs can improve modeled SWC if the soil map information is not sufficient. Temporal evolution of subsurface parameters of the jackknife simulations (e.g. jk-S80-\*) was close to the evolution of parameter estimates by other simulations (e.g. PAR-S80-10). Parameter estimates at jackknife test sites were inferred from multiple surrounding CRP sites, while updates at sites with CRP information were strongly inferred from single site information. A comparison of parameter estimates at the end of the assimilation period indicates that initial soil parameterization has a limited effect on the resulting parameter estimates. Parameter estimates of jk-BK50-30\* and jk-S80-30\* are close together at the end of the assimilation period. The CRP network led to improved results for the jackknife



evaluation simulations in case of the biased soil map. This suggests that assimilation of CRP data is particularly useful for regions with little information on subsurface parameters. We expect a tradeoff between the initial uncertainty on soil moisture content (related to the quality of the soil map and meteorological data) and the density of a CRP network. In case of a large uncertainty, like in regions with limited information about soils and a low density of meteorological stations, a sparse network of probes can already be helpful for improving soil moisture characterization. On the other hand, in regions with a high density of meteorological stations and a high resolution soil map it can be expected that a high resolution CRP network is needed to further lower the error of soil moisture characterization. Further experiments in other regions with networks of CRPs are needed to get more quantitative information about this.

A question that remains to be answered is whether it is more beneficial to assimilate neutron counts measured by CRPs directly or to assimilate CRP SWC retrievals derived from the neutron counts, as done in this study. Fast neutron intensity measured by CRPs is also affected by vegetation. Neutron count rate decreases with increasing biomass because of the hydrogen content in vegetation (Baatz et al., 2015). Seasonal biomass changes at a single site have a rather small impact on neutron intensity compared to differences between grass land site and a forest site (Baatz et al., 2015). Therefore, using measured neutron flux directly in a data assimilation framework in a catchment with different vegetation types would require to account for the effects of vegetation types on neutron intensity. Hence, vegetation estimates for each grid cell would be necessary. At present, there are two methods that include biomass in the CRP calibration process (Baatz et al., 2015; Franz et al., 2013b) but both methods naturally require accurate biomass estimates, which are typically not available. Besides the uncertainty associated with CRP methods using biomass in the calibration process, biomass estimates also come along with high uncertainties. Therefore, in the case of a catchment with different vegetation types, it is desirable to circumvent the use of biomass estimates, and assimilate directly SWC retrievals obtained at the observation sites instead of assimilating neutron intensity. Therefore, this study uses CRP SWC retrievals in the data assimilation scheme assuming that seasonal changes of biomass can be neglected.

## 6 Conclusions and Outlook

This study demonstrates the benefits of assimilating data from a network of nine cosmic-ray probes (CRP) in the land surface model CLM version 4.5. Although information on neutron flux intensity was only available at few locations in the catchment, the local ensemble transform Kalman filter (LETKF) allows updating of soil water content (SWC) at unmonitored locations in the catchment considering model and observation uncertainties. Joint state-parameter estimates improved soil moisture estimates during the assimilation and during the evaluation period. The  $E_{\text{RMS}}$  and bias for the soil moisture characterization reduced strongly for simulations initialized with a biased soil map and approached values similar to the ones obtained when the regional soil map was used as input to the simulations.  $E_{\text{RMS}}$ -values in simulations with a regional soil map were not improved, because open loop simulation results were already close to the observations. The beneficial results of joint state-parameter updates were confirmed by additional jackknife experiments. This real-world case



study on assimilating CRP SWC retrievals into a land surface model shows the potential of CRP networks to improve subsurface parameterization in regional land surface models, especially if prior information on soil properties is limited. In many areas of the world, less detailed soil maps are available than the high resolution regional soil map applied in this study. In these areas, more advanced sub-surface characterization is possible using CRP measurements and the data assimilation framework presented in this study.

For now, CRP neutron intensity observations were not assimilated directly. In future studies it would be desirable to use the COSMIC operator for assimilating neutron intensity observations directly. However, in this case the impact of biomass on the CRP measurement signal would have to be taken into account. Therefore, it is desirable to further develop the COSMIC operator to include the impact of biomass on neutron intensities. Using the biogeochemical module of CLM would then allow to characterize local vegetation states as input for the measurement operator. Remotely sensed vegetation states are another option to characterize vegetation states as input for the measurement operator. Both methods require additional field measurements for the verification of vegetation state estimates. The further extension of the data assimilation framework would also enable the estimation of additional sub-surface parameters. The impact of other sub-surface parameters such as subsurface drainage parameters and the surface drainage decay factor on SWC states and radiative surface fluxes has already been shown (Sun et al., 2013). Estimation of these parameters is desirable because of the inherent uncertainty of these globally tuned parameters. However, estimation of soil texture and organic matter content was demonstrated to be already beneficial for improved SWC modeling. Hence, this study represents a way forward towards the integration of CRP information in the calibration of large scale weather prediction models.

## 20 Data Availability

Most data presented in this study are freely available via the TERENO data portal TEODOOR (<http://teodoor.icg.kfa-juelich.de/>). Atmospheric data were licensed by the German Weather Service (DWD), and the BK50 soil map was licensed by the Geologischer Dienst Nordrhein-Westfalen.

## Acknowledgements

25 We gratefully acknowledge the support by the SFB-TR32 "Pattern in Soil-Vegetation-Atmosphere Systems: Monitoring, Modelling and Data Assimilation" funded by the Deutsche Forschungsgemeinschaft (DFG) and TERENO (Terrestrial Environmental Observatories) funded by the Helmholtz-Gemeinschaft. The authors also gratefully acknowledge the computing time granted by the John von Neumann Institute for Computing (NIC) and provided on the supercomputer JURECA at Jülich Supercomputing Centre (JSC).



## 7 References

- Ajami, H., McCabe, M. F., Evans, J. P., and Stisen, S.: Assessing the impact of model spin-up on surface water-groundwater interactions using an integrated hydrologic model, *Water Resour Res*, 50, 2636-2656, 10.1002/2013wr014258, 2014.
- Anderson, J. L.: An ensemble adjustment Kalman filter for data assimilation, *Mon Weather Rev*, 129, 2884-2903, Doi 10.1175/1520-0493(2001)129<2884: Aeakff>2.0.Co;2, 2001.
- 5 Baatz, R., Bogen, H. R., Hendricks Franssen, H. J., Huisman, J. A., Qu, W., Montzka, C., and Vereecken, H.: Calibration of a catchment scale cosmic-ray probe network: A comparison of three parameterization methods, *J Hydrol*, 516, 231-244, <http://dx.doi.org/10.1016/j.jhydrol.2014.02.026>, 2014.
- Baatz, R., Bogen, H. R., Hendricks Franssen, H. J., Huisman, J. A., Montzka, C., and Vereecken, H.: An empirical vegetation correction for soil water content quantification using cosmic ray probes, *Water Resour Res*, 51, 2030-2046, 10.1002/2014WR016443, 2015.
- 10 Bateni, S. M., and Entekhabi, D.: Surface heat flux estimation with the ensemble Kalman smoother: Joint estimation of state and parameters, *Water Resour Res*, 48, Artn W08521, 10.1029/2011wr011542, 2012.
- Bell, J. E., Leeper, R. D., Palecki, M. A., Coopersmith, E., Wilson, T., Bilotta, R., and Embler, S.: Evaluation of the 2012 Drought with a Newly Established National Soil Monitoring Network, *Vadose Zone J*, 14, 10.2136/vzj2015.02.0023, 2015.
- 15 Bogen, H. R., Herbst, M., Hake, J. F., Kunkel, R., Montzka, C., Pütz, T., Vereecken, H., and Wendland, F.: MOSYRUR - Water balance analysis in the Rur basin., in: *Schriften des Forschungszentrums Jülich. Reihe Umwelt/Environment*, Jülich, 2005.
- Bogen, H. R., Huisman, J. A., Baatz, R., Franssen, H. J. H., and Vereecken, H.: Accuracy of the cosmic-ray soil water content probe in humid forest ecosystems: The worst case scenario, *Water Resour Res*, 49, 5778-5791, Doi 10.1002/Wrcr.20463, 2013.
- 20 Burgers, G., van Leeuwen, P. J., and Evensen, G.: Analysis scheme in the ensemble Kalman filter, *Mon Weather Rev*, 126, 1719-1724, Doi 10.1175/1520-0493(1998)126<1719:Asitek>2.0.Co;2, 1998.
- Chen, F., Manning, K. W., LeMone, M. A., Trier, S. B., Alfieri, J. G., Roberts, R., Tewari, M., Niyogi, D., Horst, T. W., Oncley, S. P., Basara, J. B., and Blanken, P. D.: Description and evaluation of the characteristics of the NCAR high-resolution land data assimilation system, *J Appl Meteorol Clim*, 46, 694-713, 10.1175/Jam2463.1, 2007.
- 25 Chen, Y., and Zhang, D. X.: Data assimilation for transient flow in geologic formations via ensemble Kalman filter, *Adv Water Resour*, 29, 1107-1122, 10.1016/j.advwatres.2005.09.007, 2006.
- Clapp, R. B., and Hornberger, G. M.: Empirical Equations for Some Soil Hydraulic-Properties, *Water Resour Res*, 14, 601-604, Doi 10.1029/Wr014i004p00601, 1978.
- Cosby, B. J., Hornberger, G. M., Clapp, R. B., and Ginn, T. R.: A Statistical Exploration of the Relationships of Soil-Moisture Characteristics to the Physical-Properties of Soils, *Water Resour Res*, 20, 682-690, Doi 10.1029/Wr020i006p00682, 1984.
- 30 Cosh, M. H., Ochsner, T. E., McKee, L., Dong, J. N., Basara, J. B., Evett, S. R., Hatch, C. E., Small, E. E., Steele-Dunne, S. C., Zreda, M., and Sayde, C.: The Soil Moisture Active Passive Marena, Oklahoma, In Situ Sensor Testbed (SMAP-MOISST): Testbed Design and Evaluation of In Situ Sensors, *Vadose Zone J*, 15, 10.2136/vzj2015.09.0122, 2016.
- Crow, W. T.: Correcting land surface model predictions for the impact of temporally sparse rainfall rate measurements using an ensemble Kalman filter and surface brightness temperature observations, *J Hydrometeorol*, 4, 960-973, 2003.
- 35 Crow, W. T., Berg, A. A., Cosh, M. H., Loew, A., Mohanty, B. P., Panciera, R., de Rosnay, P., Ryu, D., and Walker, J. P.: Upscaling Sparse Ground-Based Soil Moisture Observations for the Validation of Coarse-Resolution Satellite Soil Moisture Products, *Rev Geophys*, 50, Artn Rg2002, Doi 10.1029/2011rg000372, 2012.
- 40 De Lannoy, G. J. M., Reichle, R. H., Arsenault, K. R., Houser, P. R., Kumar, S., Verhoest, N. E. C., and Pauwels, V. R. N.: Multiscale assimilation of Advanced Microwave Scanning Radiometer-EOS snow water equivalent and Moderate Resolution Imaging Spectroradiometer snow cover fraction observations in northern Colorado, *Water Resour Res*, 48, Artn W01522, Doi 10.1029/2011wr010588, 2012.
- Desilets, D., Zreda, M., and Ferre, T. P. A.: Nature's neutron probe: Land surface hydrology at an elusive scale with cosmic rays, *Water Resour Res*, 46, 10.1029/2009WR008726, 2010.
- 45 Desilets, D., and Zreda, M.: Footprint diameter for a cosmic-ray soil moisture probe: Theory and Monte Carlo simulations, *Water Resour Res*, 49, 3566-3575, Doi 10.1002/Wrcr.20187, 2013.
- Draper, C. S., Mahfouf, J. F., and Walker, J. P.: An EKF assimilation of AMSR-E soil moisture into the ISBA land surface scheme, *J Geophys Res-Atmos*, 114, 10.1029/2008JD011650, 2009.
- 50 Dunne, S., and Entekhabi, D.: An ensemble-based reanalysis approach to land data assimilation, *Water Resour Res*, 41, Artn W02013, 10.1029/2004wr003449, 2005.
- Eltahir, E. A. B.: A soil moisture rainfall feedback mechanism 1. Theory and observations, *Water Resour Res*, 34, 765-776, Doi 10.1029/97wr03499, 1998.
- Erdal, D., Neuweiler, I., and Wollschläger, U.: Using a bias aware EnKF to account for unresolved structure in an unsaturated zone model, *Water Resour Res*, 50, 132-147, 10.1002/2012wr013443, 2014.
- 55





- Erdal, D., Rahman, M. A., and Neuweiler, I.: The importance of state transformations when using the ensemble Kalman filter for unsaturated flow modeling: Dealing with strong nonlinearities, *Adv Water Resour*, 86, 354-365, 10.1016/j.advwatres.2015.09.008, 2015.
- Evensen, G.: Sequential Data Assimilation with a Nonlinear Quasi-Geostrophic Model Using Monte-Carlo Methods to Forecast Error Statistics, *J Geophys Res-Oceans*, 99, 10143-10162, Doi 10.1029/94jc00572, 1994.
- 5 Franz, T. E., Zreda, M., Ferre, T. P. A., and Rosolem, R.: An assessment of the effect of horizontal soil moisture heterogeneity on the area-average measurement of cosmic-ray neutrons, *Water Resour Res*, 49, 6450-6458, Doi 10.1002/Wrcr.20530, 2013a.
- Franz, T. E., Zreda, M., Rosolem, R., and Ferre, T. P. A.: A universal calibration function for determination of soil moisture with cosmic-ray neutrons, *Hydrol Earth Syst Sc*, 17, 453-460, DOI 10.5194/hess-17-453-2013, 2013b.
- GmbH, s.: Digital Elevation Model 10 without anthropogenic landfo<sub>RMS</sub>, 2010.
- 10 Han, X., Franssen, H. J. H., Montzka, C., and Vereecken, H.: Soil moisture and soil properties estimation in the Community Land Model with synthetic brightness temperature observations, *Water Resour Res*, 50, 6081-6105, 10.1002/2013WR014586, 2014a.
- Han, X., Franssen, H. J. H., Rosolem, R., Jin, R., Li, X., and Vereecken, H.: Correction of systematic model forcing bias of CLM using assimilation of cosmic-ray Neutrons and land surface temperature: a study in the Heihe Catchment, China, *Hydrol. Earth Syst. Sci.*, 19, 615-629, 10.5194/hess-19-615-2015, 2015.
- 15 Han, X., Franssen, H.-J. H., Bello, M. Á. J., Rosolem, R., Bogen, H., Alzamora, F. M., Chanzy, A., and Vereecken, H.: Simultaneous Soil Moisture and Properties Estimation for a Drip Irrigated Field by Assimilating Cosmic-ray Neutron Intensity, *J Hydrol*, <http://dx.doi.org/10.1016/j.jhydrol.2016.05.050>, 2016.
- Han, X. J., Franssen, H. J. H., Li, X., Zhang, Y. L., Montzka, C., and Vereecken, H.: Joint Assimilation of Surface Temperature and L-Band Microwave Brightness Temperature in Land Data Assimilation, *Vadose Zone J*, 12, Doi 10.2136/Vzj2012.0072, 2013.
- 20 Han, X. J., Franssen, H. J. H., Montzka, C., and Vereecken, H.: Soil moisture and soil properties estimation in the Community Land Model with synthetic brightness temperature observations, *Water Resour Res*, 50, 6081-6105, 10.1002/2013WR014586, 2014b.
- Hawdon, A., McJannet, D., and Wallace, J.: Calibration and correction procedures for cosmic-ray neutron soil moisture probes located across Australia, *Water Resour Res*, 50, 5029-5043, 10.1002/2013WR015138, 2014.
- Hendricks Franssen, H. J., and Kinzelbach, W.: Real-time groundwater flow modeling with the Ensemble Kalman Filter: Joint estimation of states and parameters and the filter inbreeding problem, *Water Resour Res*, 44, Artn W09408 Doi 10.1029/2007wr006505, 2008.
- 25 Houser, P. R., Shuttleworth, W. J., Famiglietti, J. S., Gupta, H. V., Syed, K. H., and Goodrich, D. C.: Integration of soil moisture remote sensing and hydrologic modeling using data assimilation, *Water Resour Res*, 34, 3405-3420, Doi 10.1029/1998wr900001, 1998.
- Hunt, B. R., Kostelich, E. J., and Szunyogh, I.: Efficient data assimilation for spatiotemporal chaos: A local ensemble transform Kalman filter, *Physica D*, 230, 112-126, DOI 10.1016/j.physd.2006.11.008, 2007.
- 30 Hurkmans, R., Paniconi, C., and Troch, P. A.: Numerical assessment of a dynamical relaxation data assimilation scheme for a catchment hydrological model, *Hydrol Process*, 20, 549-563, 10.1002/hyp.5921, 2006.
- Jackson, T. J., Le Vine, D. M., Hsu, A. Y., Oldak, A., Starks, P. J., Swift, C. T., Isham, J. D., and Haken, M.: Soil moisture mapping at regional scales using microwave radiometry: The Southern Great Plains Hydrology Experiment, *Ieee T Geosci Remote*, 37, 2136-2151, Doi 10.1109/36.789610, 1999.
- 35 Kohli, M., Schron, M., Zreda, M., Schmidt, U., Dietrich, P., and Zacharias, S.: Footprint characteristics revised for field-scale soil moisture monitoring with cosmic-ray neutrons, *Water Resour Res*, 51, 5772-5790, 10.1002/2015WR017169, 2015.
- Koster, R. D., Dirmeyer, P. A., Guo, Z. C., Bonan, G., Chan, E., Cox, P., Gordon, C. T., Kanae, S., Kowalczyk, E., Lawrence, D., Liu, P., Lu, C. H., Malyshev, S., McAvaney, B., Mitchell, K., Mocko, D., Oki, T., Oleson, K., Pitman, A., Sud, Y. C., Taylor, C. M., Versegny, D., Vasic, R., Xue, Y. K., Yamada, T., and Team, G.: Regions of strong coupling between soil moisture and precipitation, *Science*, 305, 1138-1140, DOI 10.1126/science.1100217, 2004.
- 40 Kumar, S. V., Reichle, R. H., Harrison, K. W., Peters-Lidard, C. D., Yatheendradas, S., and Santanello, J. A.: A comparison of methods for a priori bias correction in soil moisture data assimilation, *Water Resour Res*, 48, Artn W03515 Doi 10.1029/2010wr010261, 2012.
- 45 Kurtz, W., He, G. W., Kollet, S. J., Maxwell, R. M., Vereecken, H., and Franssen, H. J. H.: TerrSysMP-PDAF version 1.0): a modular high-performance data assimilation framework for an integrated land surface-subsurface model, *Geosci Model Dev*, 9, 1341-1360, 10.5194/gmd-9-1341-2016, 2016.
- Lawrence, D. M., and Slater, A. G.: Incorporating organic soil into a global climate model, *Clim Dynam*, 30, 145-160, 10.1007/s00382-007-0278-1, 2008.
- 50 Lee, J. H.: Spatial-Scale Prediction of the SVAT Soil Hydraulic Variables Characterizing Stratified Soils on the Tibetan Plateau from an EnKF Analysis of SAR Soil Moisture, *Vadose Zone J*, 13, 10.2136/vzj2014.06.0060, 2014.
- Liu, J. J., Fung, I., Kalnay, E., Kang, J. S., Olsen, E. T., and Chen, L.: Simultaneous assimilation of AIRS Xco(2) and meteorological observations in a carbon climate model with an ensemble Kalman filter, *J Geophys Res-Atmos*, 117, Artn D05309 10.1029/2011jd016642, 2012.
- 55 Lv, L., Franz, T. E., Robinson, D. A., and Jones, S. B.: Measured and Modeled Soil Moisture Compared with Cosmic-Ray Neutron Probe Estimates in a Mixed Forest, *Vadose Zone J*, 13, 10.2136/vzj2014.06.0077, 2014.



- Miyoshi, T., and Kunii, M.: The Local Ensemble Transform Kalman Filter with the Weather Research and Forecasting Model: Experiments with Real Observations, *Pure Appl Geophys*, 169, 321-333, 10.1007/s00024-011-0373-4, 2012.
- Montzka, C., Canty, M., Kunkel, R., Menz, G., Vereecken, H., and Wendland, F.: Modelling the water balance of a mesoscale catchment basin using remotely sensed land cover data, *J Hydrol*, 353, 322-334, DOI 10.1016/j.jhydrol.2008.02.018, 2008.
- 5 Montzka, C., Moradkhani, H., Weihermuller, L., Franssen, H. J. H., Canty, M., and Vereecken, H.: Hydraulic parameter estimation by remotely-sensed top soil moisture observations with the particle filter, *J Hydrol*, 399, 410-421, 10.1016/j.jhydrol.2011.01.020, 2011.
- Montzka, C., Pauwels, V. R. N., Franssen, H. J. H., Han, X. J., and Vereecken, H.: Multivariate and Multiscale Data Assimilation in Terrestrial Systems: A Review, *Sensors-Basel*, 12, 16291-16333, 10.3390/s121216291, 2012.
- Montzka, C., Grant, J. P., Moradkhani, H., Franssen, H. J. H., Weihermuller, L., Drusch, M., and Vereecken, H.: Estimation of Radiative Transfer Parameters from L-Band Passive Microwave Brightness Temperatures Using Advanced Data Assimilation, *Vadose Zone J*, 12, 10.2136/vzj2012.0040, 2013.
- 10 Moradkhani, H., Sorooshian, S., Gupta, H. V., and Houser, P. R.: Dual state-parameter estimation of hydrological models using ensemble Kalman filter, *Adv Water Resour*, 28, 135-147, 10.1016/j.advwatres.2004.09.002, 2005.
- Nordrhein-Westfalen, G. D.: Informationssystem Bodenkarte 50, 1:50000, 2009.
- 15 Oglesby, R. J., and Erickson, D. J.: Soil-Moisture and the Persistence of North-American Drought, *J Climate*, 2, 1362-1380, Doi 10.1175/1520-0442(1989)002<1362:Smato>2.0.Co;2, 1989.
- Oleson, K., Lawrence, D. M., Bonan, G. B., Drewniak, B., Huang, M., Koven, C. D., Levis, S., Li, F., Riley, J. M., Subin, Z. M., Swenson, S., Thornton, P. E., Bozbiyik, A., Fisher, R., Heald, C. L., Kluzek, E., Lamarque, J.-F., Lawrence, P. J., Leung, L. R., Lipscomb, W., Muszala, S. P., Ricciuto, D. M., Sacks, W. J., Sun, Y., Tang, J., and Yang, Z.-L.: Technical description of version 4.5 of the Community Land Model (CLM), NCAR Technical Note NCAR/TN-503+STR, 420, 10.5065/D6RR1W7M, 2013.
- 20 Pauwels, V. R. N., Balenzano, A., Satalino, G., Skriver, H., Verhoest, N. E. C., and Mattia, F.: Optimization of Soil Hydraulic Model Parameters Using Synthetic Aperture Radar Data: An Integrated Multidisciplinary Approach, *Ieee T Geosci Remote*, 47, 455-467, 10.1109/Tgrs.2008.2007849, 2009.
- Pelowitz, D. B.: MCNPX user's manual, version 5, Rep. LA-CP-05-0369, Los Alamos National Laboratory, Los Alamos LA-CP-05-0369, 25 2005.
- Penny, S. G., Kalnay, E., Carton, J. A., Hunt, B. R., Ide, K., Miyoshi, T., and Chepurin, G. A.: The local ensemble transform Kalman filter and the running-in-place algorithm applied to a global ocean general circulation model, *Nonlinear Proc Geoph*, 20, 1031-1046, 10.5194/npg-20-1031-2013, 2013.
- Reichle, R. H., McLaughlin, D. B., and Entekhabi, D.: Hydrologic data assimilation with the ensemble Kalman filter, *Mon Weather Rev*, 130, 103-114, Doi 10.1175/1520-0493(2002)130<0103:Hdawte>2.0.Co;2, 2002a.
- 30 Reichle, R. H., Walker, J. P., Koster, R. D., and Houser, P. R.: Extended versus ensemble Kalman filtering for land data assimilation, *J Hydrometeorol*, 3, 728-740, Doi 10.1175/1525-7541(2002)003<0728:Evekf>2.0.Co;2, 2002b.
- Reichle, R. H.: Data assimilation methods in the Earth sciences, *Adv Water Resour*, 31, 1411-1418, 10.1016/j.advwatres.2008.01.001, 2008.
- 35 Reichle, R. H., Kumar, S. V., Mahanama, S. P. P., Koster, R. D., and Liu, Q.: Assimilation of Satellite-Derived Skin Temperature Observations into Land Surface Models, *J Hydrometeorol*, 11, 1103-1122, Doi 10.1175/2010jhm1262.1, 2010.
- Rhodin, A., Kucharski, F., Callies, U., Eppel, D. P., and Wergen, W.: Variational analysis of effective soil moisture from screen-level atmospheric parameters: Application to a short-range weather forecast model, *Q J Roy Meteor Soc*, 125, 2427-2448, Doi 10.1256/Smsqj.55904, 1999.
- 40 Robinson, D. A., Campbell, C. S., Hopmans, J. W., Hornbuckle, B. K., Jones, S. B., Knight, R., Ogden, F., Selker, J., and Wendroth, O.: Soil moisture measurement for ecological and hydrological watershed-scale observatories: A review, *Vadose Zone J*, 7, 358-389, Doi 10.2136/Vzj2007.0143, 2008.
- Rosolem, R., Hoar, T., Arellano, A., Anderson, J. L., Shuttleworth, W. J., Zeng, X., and Franz, T. E.: Translating aboveground cosmic-ray neutron intensity to high-frequency soil moisture profiles at sub-kilometer scale, *Hydrol. Earth Syst. Sci.*, 18, 4363-4379, 10.5194/hess-18-4363-2014, 2014.
- 45 Seneviratne, S. I., Luthi, D., Litschi, M., and Schar, C.: Land-atmosphere coupling and climate change in Europe, *Nature*, 443, 205-209, Doi 10.1038/Nature05095, 2006.
- Seneviratne, S. I., Corti, T., Davin, E. L., Hirschi, M., Jaeger, E. B., Lehner, I., Orlowsky, B., and Teuling, A. J.: Investigating soil moisture-climate interactions in a changing climate: A review, *Earth-Science Reviews*, 99, 125-161, http://dx.doi.org/10.1016/j.earscirev.2010.02.004, 2010.
- 50 Sheffield, J., and Wood, E. F.: Global trends and variability in soil moisture and drought characteristics, 1950-2000, from observation-driven Simulations of the terrestrial hydrologic cycle, *J Climate*, 21, 432-458, 10.1175/2007JCLI1822.1, 2008.
- Shi, Y. N., Davis, K. J., Zhang, F. Q., Duffy, C. J., and Yu, X.: Parameter estimation of a physically based land surface hydrologic model using the ensemble Kalman filter : A synthetic experiment, *Water Resour Res*, 50, 706-724, 10.1002/2013wr014070, 2014.





- Shi, Y. N., Davis, K. J., Zhang, F. Q., Duffy, C. J., and Yu, X.: Parameter estimation of a physically-based land surface hydrologic model using an ensemble Kalman filter: A multivariate real-data experiment, *Adv Water Resour*, 83, 421-427, 10.1016/j.advwatres.2015.06.009, 2015.
- Shrestha, P., Sulis, M., Masbou, M., Kollet, S., and Simmer, C.: A Scale-Consistent Terrestrial Systems Modeling Platform Based on COSMO, CLM, and ParFlow, *Mon Weather Rev*, 142, 3466-3483, 10.1175/Mwr-D-14-00029.1, 2014.
- 5 Shuttleworth, J., Rosolem, R., Zreda, M., and Franz, T.: The COsmic-ray Soil Moisture Interaction Code (COSMIC) for use in data assimilation, *Hydrol Earth Syst Sc*, 17, 3205-3217, DOI 10.5194/hess-17-3205-2013, 2013.
- Simmer, C., Thiele-Eich, I., Masbou, M., Amelung, W., Bogen, H., Crewell, S., Diekkruiger, B., Ewert, F., Franssen, H. J. H., Huisman, J. A., Kemna, A., Klitzsch, N., Kollet, S., Langensiepen, M., Lohnert, U., Rahman, A. S. M. M., Rascher, U., Schneider, K., Schween, J., Shao, Y. P., Shrestha, P., Stiebler, M., Sulis, M., Vanderborght, J., Vereecken, H., van der Kruk, J., Waldhoff, G., and Zerenner, T.: MONITORING AND MODELING THE TERRESTRIAL SYSTEM FROM PORES TO CATCHMENTS The Transregional Collaborative Research Center on Patterns in the Soil-Vegetation-Atmosphere System, *B Am Meteorol Soc*, 96, 1765-1787, 10.1175/Bams-D-13-00134.1, 2015.
- 10 Song, X. H., Shi, L. S., Ye, M., Yang, J. Z., and Navon, I. M.: Numerical Comparison of Iterative Ensemble Kalman Filters for Unsaturated Flow Inverse Modeling, *Vadose Zone J*, 13, 10.2136/vzj2013.05.0083, 2014.
- 15 Sun, Y., Hou, Z., Huang, M., Tian, F., and Leung, L. R.: Inverse modeling of hydrologic parameters using surface flux and runoff observations in the Community Land Model, *Hydrol Earth Syst Sc*, 17, 4995-5011, 10.5194/hess-17-4995-2013, 2013.
- Temimi, M., Lakhankar, T., Zhan, X. W., Cosh, M. H., Krakauer, N., Fares, A., Kelly, V., Khanbilvardi, R., and Kumassi, L.: Soil Moisture Retrieval Using Ground-Based L-Band Passive Microwave Observations in Northeastern USA, *Vadose Zone J*, 13, 10.2136/vzj2013.06.0101, 2014.
- 20 Vereecken, H., Huisman, J. A., Bogen, H., Vanderborght, J., Vrugt, J. A., and Hopmans, J. W.: On the value of soil moisture measurements in vadose zone hydrology: A review, *Water Resour Res*, 44, Artn W00d06  
Doi 10.1029/2008wr006829, 2008.
- Vereecken, H., Schnepf, A., Hopmans, J. W., Javaux, M., Or, D., Roose, D. O. T., Vanderborght, J., Young, M. H., Amelung, W., Aitkenhead, M., Allison, S. D., Assouline, S., Baveye, P., Berli, M., Bruggemann, N., Finke, P., Flury, M., Gaiser, T., Govers, G., Ghezzehei, T., Hallett, P., Franssen, H. J. H., Heppell, J., Horn, R., Huisman, J. A., Jacques, D., Jonard, F., Kollet, S., Lafolie, F., Lamorski, K., Leitner, D., McBratney, A., Minasny, B., Montzka, C., Nowak, W., Pachepsky, Y., Padarian, J., Romano, N., Roth, K., Rothfuss, Y., Rowe, E. C., Schwen, A., Simunek, J., Tiktak, A., Van Dam, J., van der Zee, S. E. A. T. M., Vogel, H. J., Vrugt, J. A., Wohling, T., and Young, I. M.: Modeling Soil Processes: Review, Key Challenges, and New Perspectives, *Vadose Zone J*, 15, 10.2136/vzj2015.09.0131, 2016.
- 30 Villarreces, C. A. R., Baroni, G., and Oswald, S. E.: Inverse modelling of cosmic-ray soil moisture for field-scale soil hydraulic parameters, *Eur J Soil Sci*, 65, 876-886, 10.1111/ejss.12162, 2014.
- Vrugt, J. A., Diks, C. G. H., Gupta, H. V., Bouten, W., and Verstraten, J. M.: Improved treatment of uncertainty in hydrologic modeling: Combining the strengths of global optimization and data assimilation, *Water Resour Res*, 41, Artn W01017  
10.1029/2004wr003059, 2005.
- 35 Waldhoff, G.: Enhanced land use classification of 2009 for the Rur catchment, in, TR32DB, 2012.
- Whan, K., Zscheischler, J., Orth, R., Shongwe, M., Rahimi, M., Asare, E. O., and Seneviratne, S. I.: Impact of soil moisture on extreme maximum temperatures in Europe, *Weather and Climate Extremes*, 9, 57-67, <http://dx.doi.org/10.1016/j.wace.2015.05.001>, 2015.
- Wu, C. C., and Margulis, S. A.: Feasibility of real-time soil state and flux characterization for wastewater reuse using an embedded sensor network data assimilation approach, *J Hydrol*, 399, 313-325, 10.1016/j.jhydrol.2011.01.011, 2011.
- 40 Wu, C. C., and Margulis, S. A.: Real-Time Soil Moisture and Salinity Profile Estimation Using Assimilation of Embedded Sensor Datastreams, *Vadose Zone J*, 12, 10.2136/vzj2011.0176, 2013.
- Zacharias, S., Bogen, H., Samaniego, L., Mauder, M., Fuss, R., Putz, T., Frenzel, M., Schwank, M., Baessler, C., Butterbach-Bahl, K., Bens, O., Borg, E., Brauer, A., Dietrich, P., Hajnsek, I., Helle, G., Kiese, R., Kunstmann, H., Klotz, S., Munch, J. C., Papen, H., Priesack, E., Schmid, H. P., Steinbrecher, R., Rosenbaum, U., Teutsch, G., and Vereecken, H.: A Network of Terrestrial Environmental Observatories in Germany, *Vadose Zone J*, 10, 955-973, Doi 10.2136/Vzj2010.0139, 2011.
- Zreda, M., Desilets, D., Ferre, T. P. A., and Scott, R. L.: Measuring soil moisture content non-invasively at intermediate spatial scale using cosmic-ray neutrons, *Geophys Res Lett*, 35, 10.1029/2008GL035655, 2008.
- Zreda, M., Shuttleworth, W. J., Zeng, X., Zweck, C., Desilets, D., Franz, T., and Rosolem, R.: COSMOS: the COsmic-ray Soil Moisture Observing System (vol 16, pg 4079, 2012), *Hydrol Earth Syst Sc*, 17, 1065-1066, DOI 10.5194/hess-17-1065-2013, 2012.
- 50



## Figures

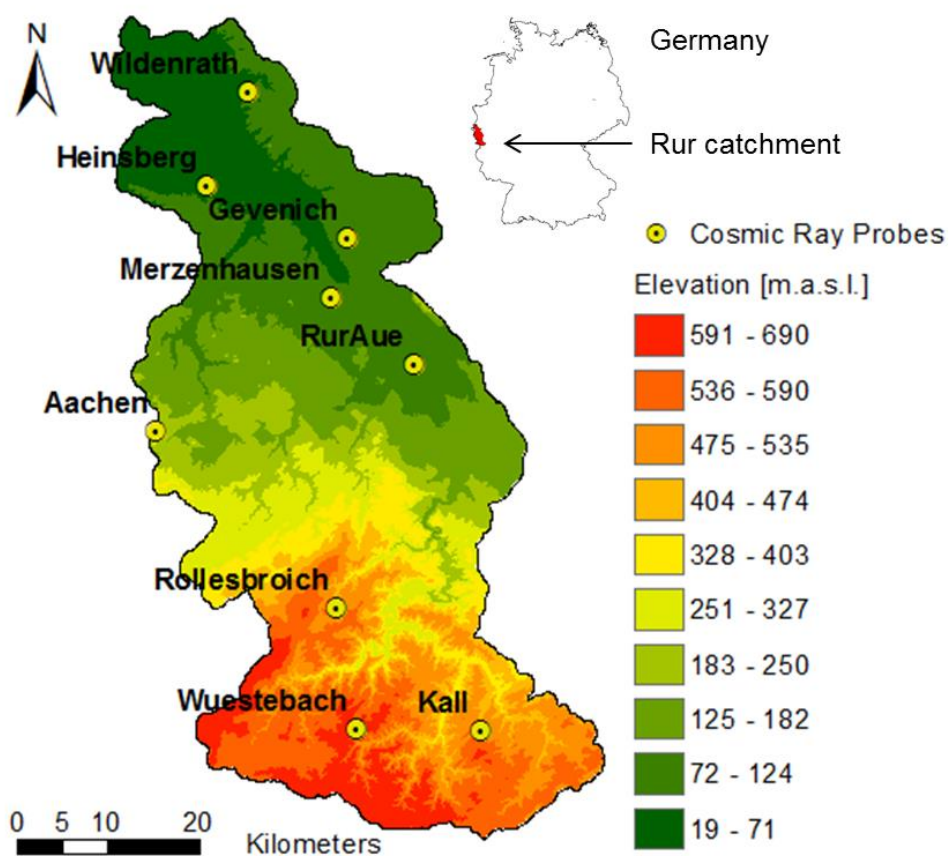


Fig. 1. Map of the Rur catchment and locations of the nine cosmic-ray probes. The hilly South of the catchment is prone to more rainfall,  
 5 lower average temperatures and less potential evapotranspiration than the North of the catchment.

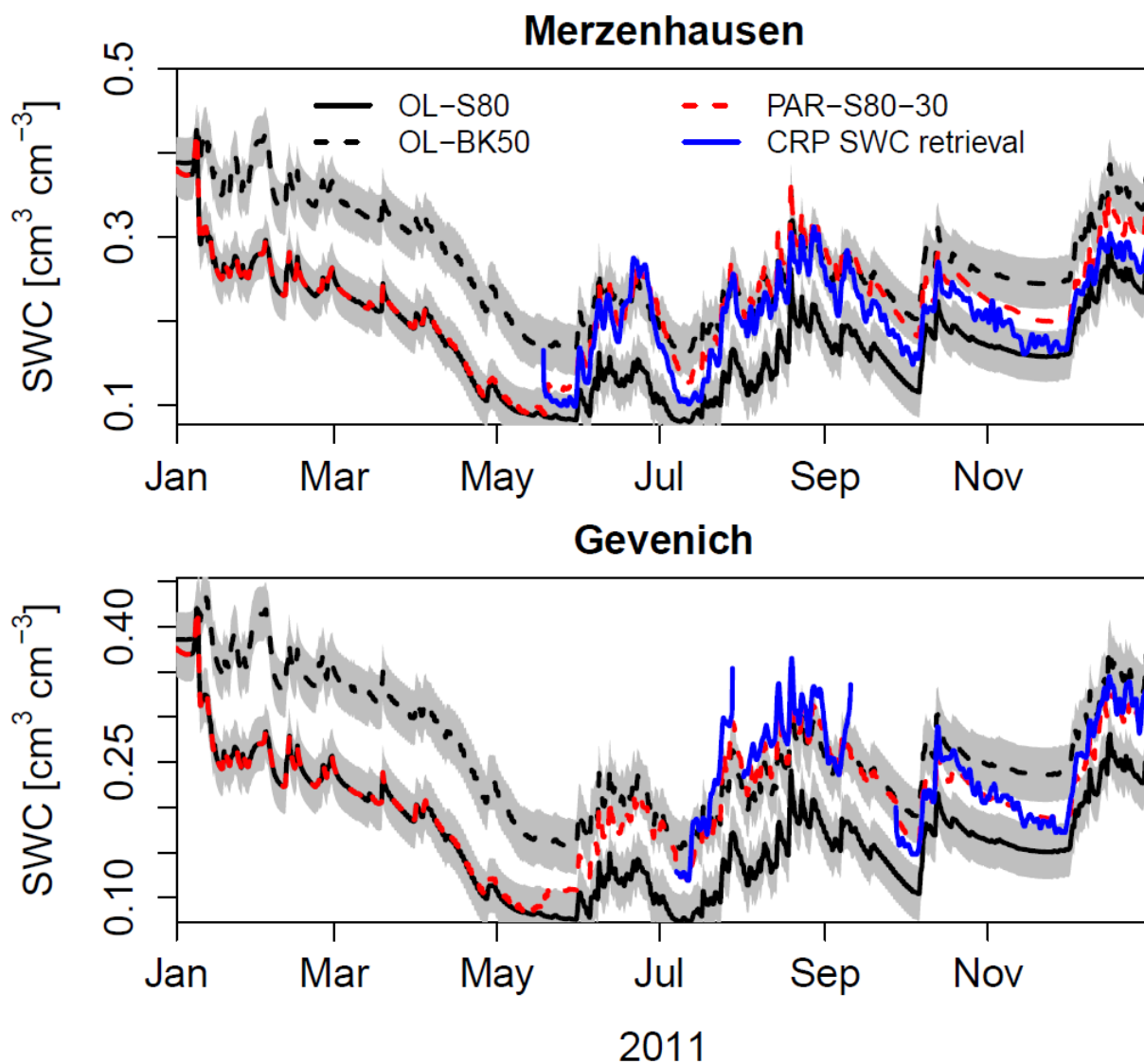


Fig. 2. Temporal evolution of simulated SWC, calculated with open loop (OL-\*) simulations and data assimilation including parameter updating (PAR-S80-30), together with the CRP soil water content retrieval (SWC) during the first year of simulation at the sites Merzenhausen and Gevenich. Simulated SWC was vertically weighted using the COSMIC operator to obtain the appropriate SWC corresponding to the CRP SWC retrieval.

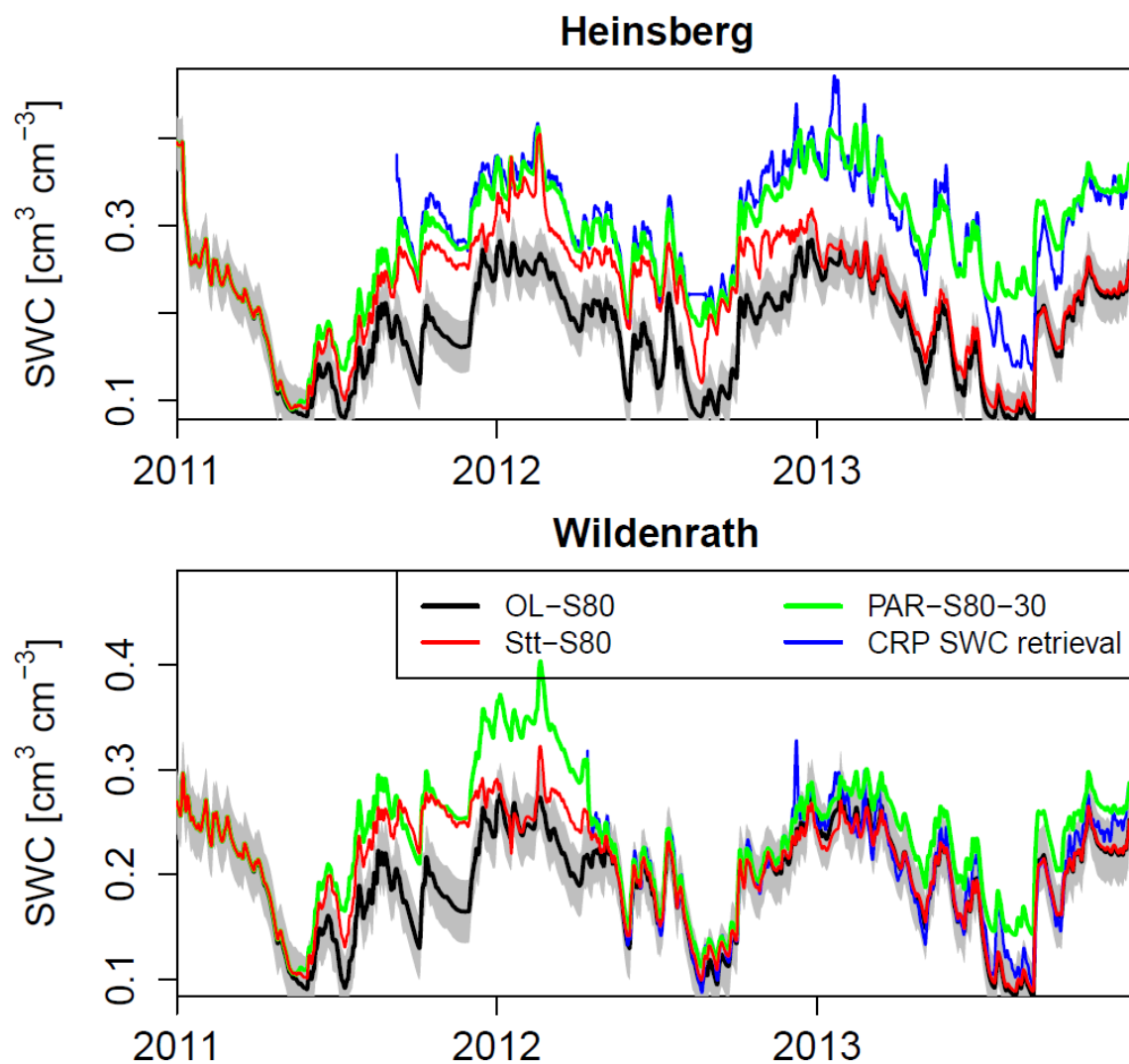


Fig. 3. Temporal evolution of simulated SWC retrievals, calculated with open loop (OL-S80), data assimilation with state update only (Stt-BK50), and data assimilation including parameter updating (PAR-S80-30), together with the CRP soil water content (SWC) retrieval at the sites Heinsberg and Wildenrath for the data assimilation period 2011 and 2012, and the evaluation period 2013. Simulated SWC was vertically weighted to obtain the appropriate SWC corresponding to the CRP SWC retrieval.

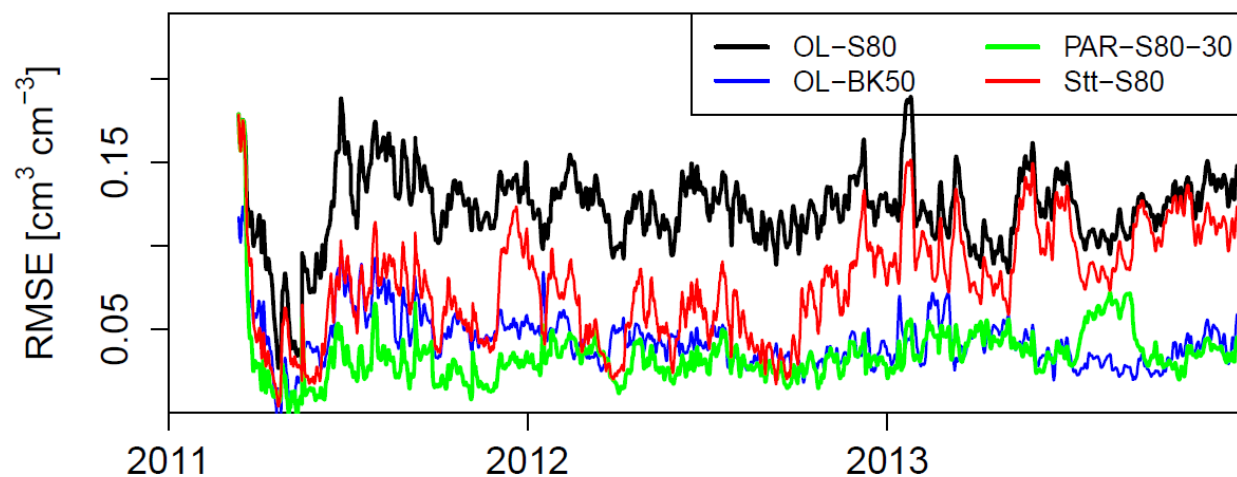


Fig. 4. Temporal evolution of root mean square error ( $E_{RMS}$ ) for hourly SWC retrievals.  $E_{RMS}$  is calculated hourly for nine CRP's for open loop runs for soil maps S80 and BK50, joint state-parameter updates (PAR-S80-30) and state updates only (Stt-S80) during the assimilation period (2011 and 2012) and verification period (2013).

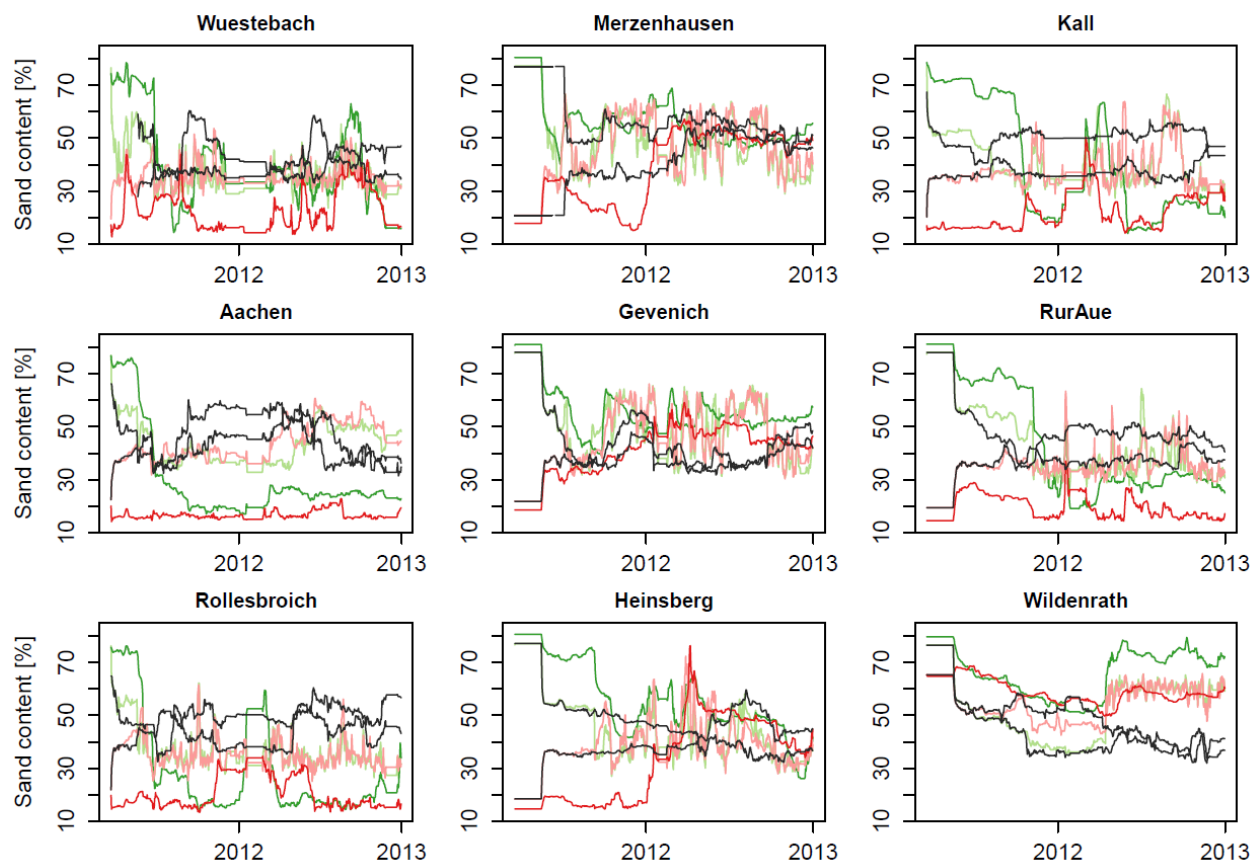


Fig. 5. Temporal evolution of the percentage sand content for simulations with parameter update: PAR-S80-30 (green), PAR-S80-10 (light green), PAR-BK50-30 (red), PAR-BK50-10 (light red), jk-S80-30\* (black) and jk-BK50-30\* (black).

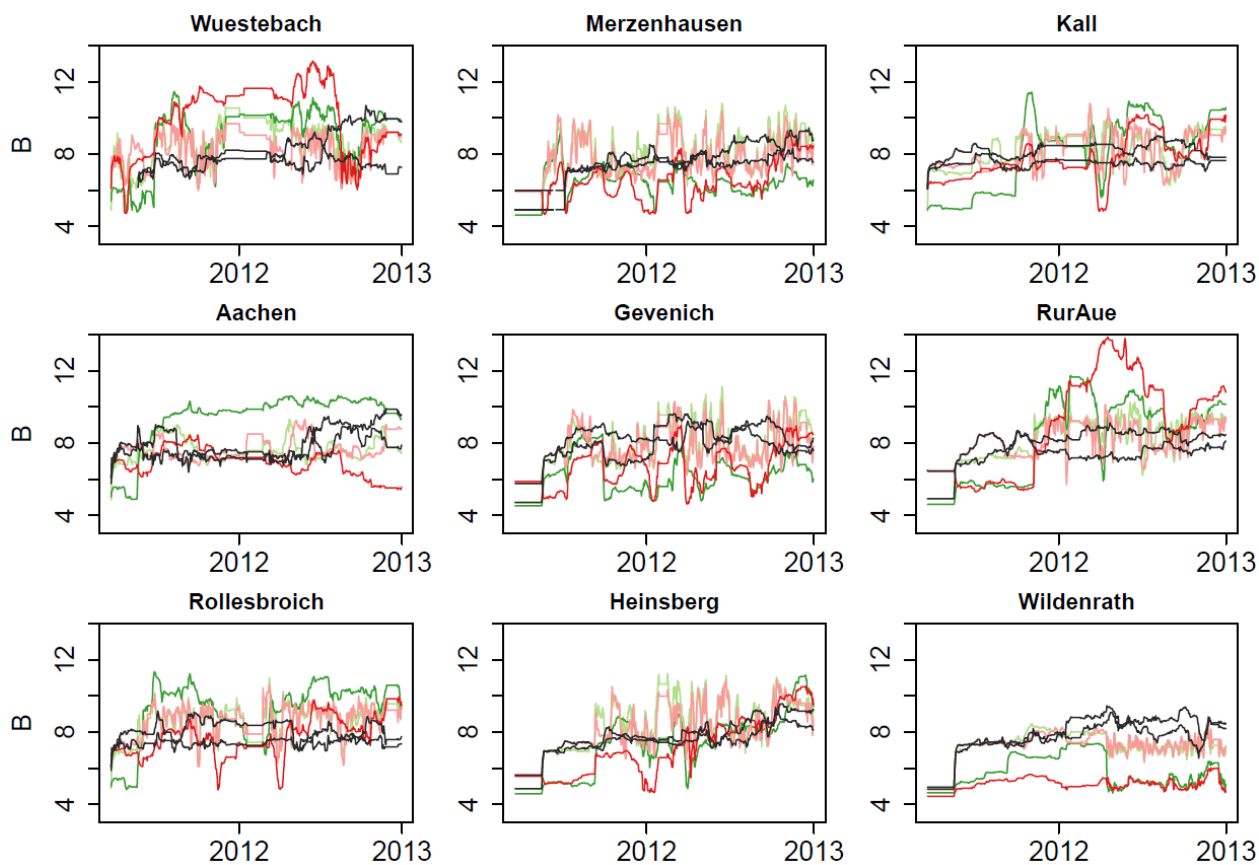


Fig. 6. Temporal evolution of the B parameter (top 15cm) for simulations with parameter update: PAR-S80-30 (green), PAR-S80-10 (light green), PAR-BK50-30 (red), PAR-BK50-10 (light red), jk-S80-30\* (black) and jk-BK50-30\* (black).



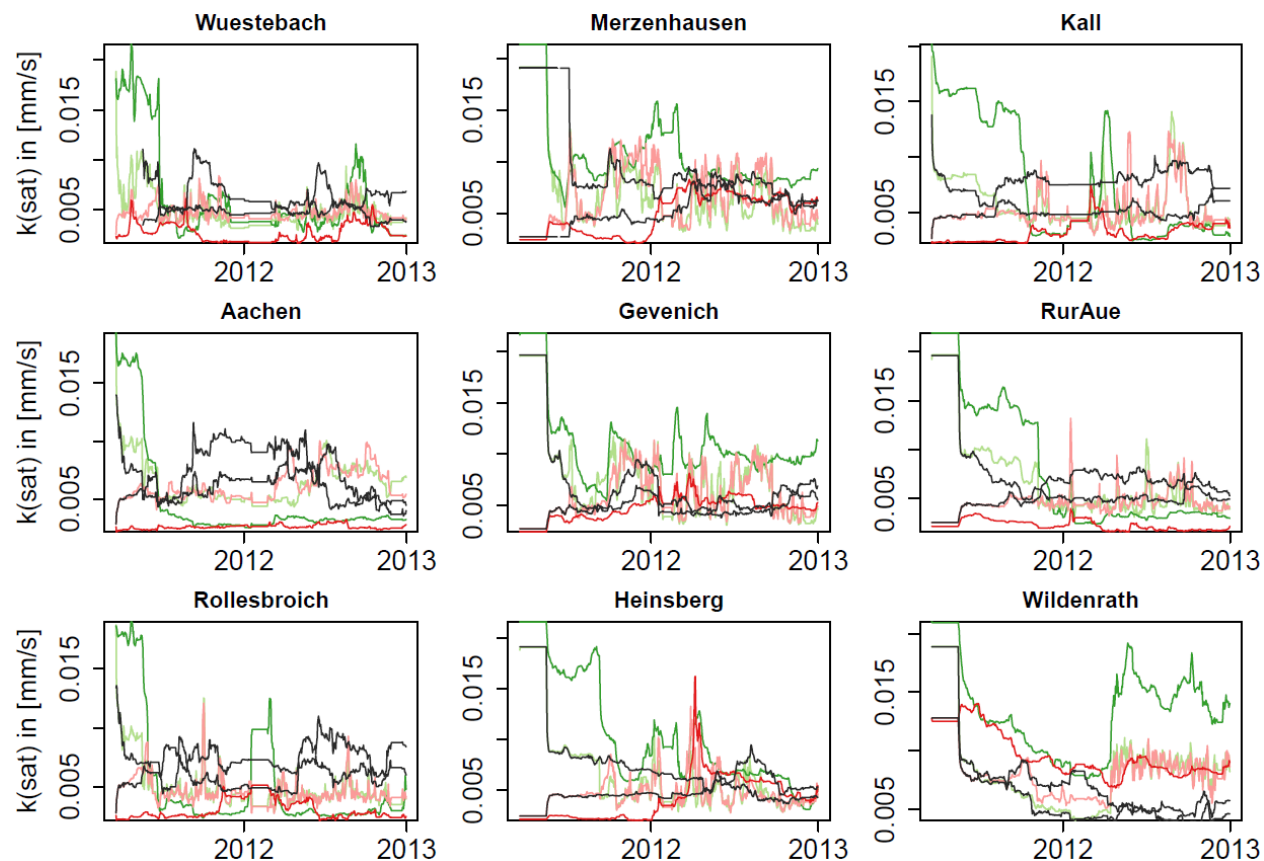


Fig. 7. Temporal evolution of saturated hydraulic conductivity (top 15cm) for simulations with parameter update: PAR-S80-30 (green), PAR-S80-10 (light green), PAR-BK50-30 (red), PAR-BK50-10 (light red), jk-S80-30\* (black) and jk-BK50-30\* (black).

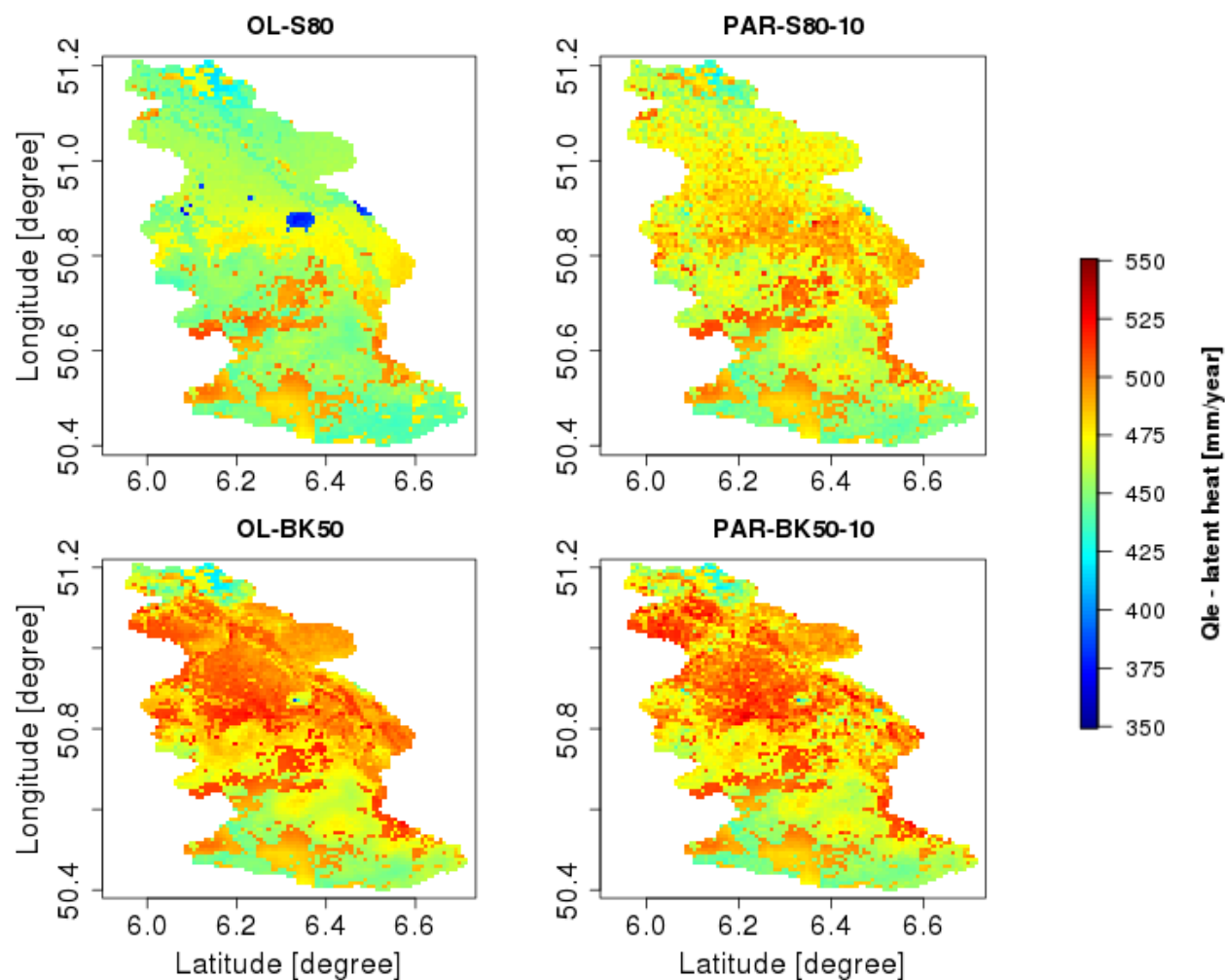


Fig. 8. Annual evapotranspiration in the evaluation period (year 2013) for simulations OL-S80, OL-BK50, PAR-S80-10 and PAR-BK50-10.



## Tables

Table 1: Site information on elevation (m.a.s.l.), average annual precipitation (mm/year), CLM plant functional type, sand content (%), clay content (%), and the date of the first SWC retrieval assimilated.

Name	m.a.s.l.	Precip.	Plant functional type	Sand	Clay	Date of first assimilation
Aachen	232	952	Crops	22	23	13.01.2012
Gevenich	108	884	Crops	22	20	07.07.2011
Heinsberg	57	814	Crops	18	19	09.09.2011
Kall	504	935	C3 non arctic grass	20	22	15.09.2011
Merzenhausen	94	825	Crops	21	22	19.05.2011
Rollesbroich	515	1307	C3 non arctic grass	22	23	19.05.2011
RurAue	102	743	C3 non arctic grass	19	26	08.11.2011
Wildenrath	76	856	Broadleaf deciduous temperate tree	65	12	07.05.2012
Wuestebach	605	1401	Needleleaf evergreen temperate tree	19	23	20.03.2011



Table 2: Overview of simulation scenarios: Open loop (OL-\*) with variation in the soil map BK50 or S80, data assimilation run with state update (Stt) or joint state- and parameter update (PAR) with variation in the soil map perturbation (-10 or -30), and jackknife evaluation runs (jk-S80-1 to 9, and jk-BK50-1 to 9).

Simulation Code	Soil Perturbation		Sand Content		Update	
	10	30	BK50	80 % fix	State	Parameter
OL-BK50		+	+			
OL-S80		+		+		
Stt-BK50		+	+		+	
Stt-S80		+		+	+	
PAR-BK50-30		+	+		+	+
PAR-BK50-10	+		+		+	+
PAR-S80-30		+		+	+	+
PAR-S80-10	+			+	+	+
jk-BK50-1 to 9		+	+		+	+
jk-S80-1 to 9		+		+	+	+



Table 3:  $E_{\text{RMS}}$  ( $\text{cm}^3/\text{cm}^3$ ) at CRP sites for open loop runs and different data assimilation scenarios, for the assimilation period (2011 and 2012).

2011 & 2012	Rolles- broich	Merzen- hausen	Geve- nich	Heins- berg	Kall	RurAue	Wueste- bach	Aachen	Wilden- rath	Average $E_{\text{RMS}}$
OL-BK50	0.054	0.067	0.039	0.035	0.042	0.027	0.041	0.032	0.017	0.039
Stt-BK50	0.033	0.041	0.021	0.022	0.030	0.024	0.038	0.023	0.017	0.028
PAR-BK50-10	0.036	0.036	0.019	0.021	0.033	0.025	0.035	0.045	0.015	0.029
PAR-BK50-30	0.031	0.034	0.018	0.019	0.027	0.023	0.040	0.044	0.016	0.028
jk-BK50-*	0.070	0.058	0.073	0.035	0.048	0.050	0.053	0.050	0.091	0.059
OL-S80	0.170	0.053	0.081	0.117	0.149	0.158	0.065	0.169	0.020	0.109
Stt-S80	0.104	0.020	0.037	0.051	0.083	0.056	0.060	0.086	0.018	0.057
PAR-S80-10	0.032	0.038	0.024	0.023	0.033	0.023	0.036	0.048	0.015	0.030
PAR-S80-30	0.029	0.035	0.018	0.019	0.027	0.023	0.039	0.068	0.016	0.030
jk-S80-*	0.082	0.038	0.063	0.026	0.062	0.034	0.038	0.073	0.095	0.057



Table 4:  $E_{\text{RMS}}$  ( $\text{cm}^3/\text{cm}^3$ ) at CRP-sites for open loop, data assimilation and jackknife simulations on the basis of a comparison with CRP SWC retrievals during the verification period (2013). For each jackknife simulation only one  $E_{\text{RMS}}$  is reported: The  $E_{\text{RMS}}$  of the location that is meant for evaluation.

Year 2013	Rolles- broich	Merzen- hausen	Geve- nich	Heins- berg	Kall	RurAue	Wueste- bach	Aachen	Wilden- rath	Average $E_{\text{RMS}}$
OL-BK50	0.044	0.065	0.036	0.027	0.048	0.038	0.048	0.042	0.017	0.041
Stt-BK50	0.041	0.054	0.034	0.027	0.049	0.038	0.048	0.041	0.018	0.039
PAR-BK50-10	0.068	0.062	0.036	0.038	0.056	0.056	0.043	0.058	0.017	0.048
PAR-BK50-30	0.052	0.061	0.035	0.033	0.068	0.048	0.043	0.048	0.035	0.047
jk-BK50-*	0.036	0.047	0.028	0.025	0.042	0.031	0.040	0.054	0.106	0.045
OL-S80	0.157	0.062	0.106	0.115	0.160	0.154	0.099	0.167	0.019	0.115
Stt-S80	0.100	0.063	0.107	0.106	0.099	0.146	0.097	0.158	0.020	0.100
PAR-S80-10	0.060	0.039	0.043	0.040	0.064	0.043	0.052	0.060	0.019	0.047
PAR-S80-30	0.049	0.059	0.037	0.036	0.053	0.032	0.046	0.047	0.035	0.044
jk-S80-*	0.079	0.046	0.042	0.036	0.059	0.038	0.063	0.044	0.105	0.057





Table 5: Bias ( $\text{cm}^3/\text{cm}^3$ ) at CRP-sites for open loop, data assimilation and jackknife simulations compared to CRP SWC retrievals during the verification period (2013). For each jackknife simulation only one bias is reported: The bias of the location that is meant for evaluation.

Year 2013	Rolles- broich	Merzen- hausen	Geve- nich	Heins- berg	Kall	Rur- Aue	Wueste- bach	Aachen	Wilden- rath	Mean absolute bias
OL-BK50	-0.03	0.06	0.01	0.00	-0.02	0.00	-0.02	0.01	0.00	0.02
Stt-BK50	-0.01	0.04	0.00	0.00	-0.01	-0.01	-0.02	0.00	0.00	0.01
PAR-BK50-10	0.06	0.05	0.01	0.02	0.04	0.04	0.02	-0.04	0.00	0.03
PAR-BK50-30	0.03	0.05	0.00	0.02	0.04	0.03	-0.01	-0.03	0.03	0.03
jk-BK50-*	-0.02	0.04	0.01	-0.01	-0.03	-0.02	-0.03	-0.05	0.11	0.04
OL-S80	-0.17	-0.05	-0.08	-0.12	-0.15	-0.16	-0.09	-0.17	-0.01	0.11
Stt-S80	-0.09	-0.05	-0.10	-0.10	-0.08	-0.14	-0.09	-0.15	-0.01	0.09
PAR-S80-10	0.03	0.05	-0.01	0.02	0.03	0.01	-0.01	-0.03	0.03	0.02
PAR-S80-30	0.04	0.02	-0.03	0.02	0.05	0.03	0.03	-0.04	-0.01	0.03
jk-S80-*	-0.07	0.03	0.02	0.02	-0.04	-0.02	-0.04	-0.03	0.10	0.04

5